CRYSTUNE: CRYSTAL GENERATION VIA FINE-TUNING OF LARGE LANGUAGE MODELS ON WYCKOFF REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

The discovery of novel materials is essential for scientific and technological progress. Recent work has explored fine-tuning large language models (LLMs) for autoregressive crystal generation, but how much crystallographic symmetry such models know, and whether it can be taught, remains unclear. We propose CrysTune, a class of LLMs fine-tuned on a Wyckoff representation of crystals with two auxiliary tasks, *Wyckoff lookup* and *canonicalization*, that teach the model the symmetry rules its representation relies on. These tasks substantially improve the model's knowledge of Wyckoff positions, which pretrained LLMs largely lack. CrysTune shows improved stability-related metrics relative to LLMs fine-tuned on standard string-encoded representations. We further use CrysTune as the initial policy for reinforcement learning fine-tuning with rewards for stability, uniqueness, novelty, and diversity. While demonstrating competitive performance of CrysTune models, we also highlight challenges in simultaneously improving multiple properties.

1 INTRODUCTION

Discovering new crystalline materials means searching a space that is combinatorial in composition and symmetry and continuous in atomic coordinates and lattice geometry. Generative models and the crystal representations designed for them have advanced this search considerably over the last five years (Xie et al., 2021; Jiao et al., 2023; Zeni et al., 2023; Miller et al., 2024). From a machine-learning perspective, progress is typically measured by whether a model produces crystals that are valid, stable, unique, and novel (SUN) and diverse, and that meet target property constraints where these are specified. Crystals have a structure that generative models can exploit: each belongs to one of 230 space groups, and every atom occupies one of a finite set of symmetry-determined Wyckoff positions. Because materials data are scarce, encoding this symmetry as an inductive bias has been a recurring theme in diffusion, flow, and transformer models (Jiao et al., 2024; Levy et al., 2025; Cao et al., 2025; Kazeev et al., 2025).

Fine-tuning pretrained large language models (LLMs), whether by supervised fine-tuning (SFT), reinforcement learning (RL), or both, is an attractive alternative to training domain-specific architectures: it needs no custom tokenizer or architecture, supports conditional generation through natural-language prompts, and benefits from the model's pretrained knowledge. Crystal-Text-LLM (Gruver et al., 2024) and CrystalLLM (Antunes et al., 2024) fine-tune LLMs on CIF-like strings of atom types, coordinates, and lattice parameters. These representations carry no explicit symmetry information, so the model must rediscover space-group constraints from coordinates alone. Wyckoff-based representations remove this burden: custom transformers such as CrystalFormer (Cao et al., 2025), WyFormer (Kazeev et al., 2025), and Matra-Genoa (De Breuck et al., 2026) generate a space group and a sequence of Wyckoff sites, and PLaID++ (Xu et al., 2025) recently fine-tuned an LLM on such strings. This raises a natural question: *how much of the Wyckoff tables does a pretrained LLM already know, and can fine-tuning supply what is missing?* A model that generates "`167 18e`" must know that this site has template $(x, 0, \frac{1}{4})$ and multiplicity 18, and the International Tables contain 1,731 such (space group, position) entries. We test Qwen-3.5 models from 0.8B to 27B parameters on this lookup task and find that even the largest recalls fewer than 30% of the templates (Section 4.1). This motivates auxiliary tasks that teach the tables explicitly, so that a model can use this knowledge to generate better crystals from a comparatively small training set. As for RL fine-tuning, rewards for

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.

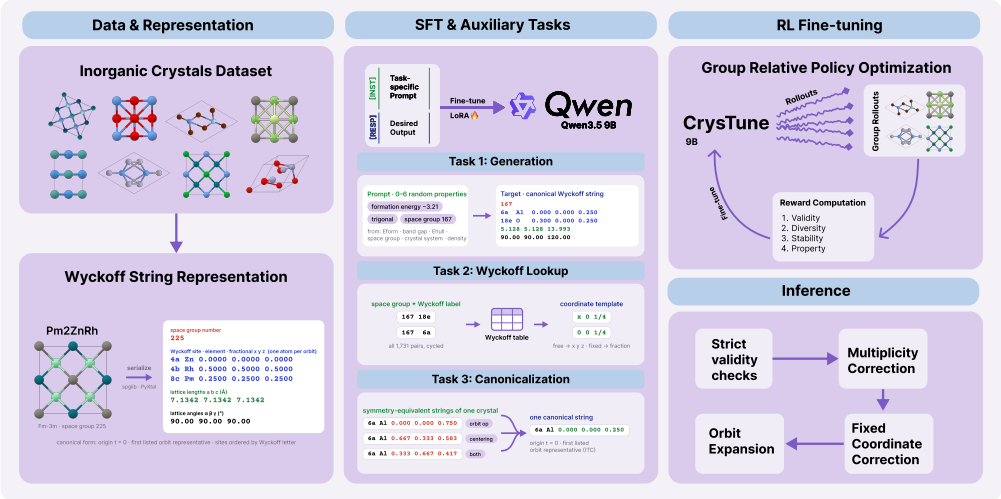


Figure 1: Overview of CrysTune. Crystals are serialized into a canonical Wyckoff string. A pretrained LLM is fine-tuned on a curriculum of three tasks: generation, Wyckoff lookup (predicting the coordinate template of a Wyckoff position), and canonicalization. The resulting policy initializes GRPO fine-tuning with stability, uniqueness, novelty, and diversity rewards. At inference, generated strings are parsed with symmetry-aware corrections before evaluation.

stability, novelty, and diversity can be computed directly from a generated structure, and have been used to steer diffusion models (Park & Walsh, 2025) and LLMs (Xu et al., 2025). We use them to align a supervised policy towards more stable crystals and towards a requested crystal system. This paper makes the following contributions.

1. **CrysTune SFT with auxiliary symmetry tasks.** We fine-tune Qwen-3.5-9B (Qwen Team, 2026) on canonical Wyckoff strings with a curriculum of two auxiliary tasks, exhaustive *Wyckoff lookup* and *canonicalization* of symmetry-equivalent inputs. We first measure how well pretrained models recall Wyckoff coordinate templates and show that fine-tuning raises the 9B model from 9.7% to 66% exact-match. Fine-tuning the same pretrained model on the same data, the Wyckoff representation yields higher stability, SUN, and space-group diversity than the CIF-style representation of Crystal-Text-LLM.
2. **Investigation of GRPO fine-tuning.** We fine-tune the SFT policy with GRPO under a reward inspired by Chemeleon2 (Park & Walsh, 2025) that combines stability, uniqueness, novelty, and diversity. To encourage exploration, we maintain a buffer so that uniqueness is measured across steps rather than within a batch, expand the novelty reference to 846k structures, and add an entropy bonus. The resulting policies improve validity and yield more metastable structures that are also unique and novel, while retaining the novelty of the supervised model, and generate a higher proportion of crystals matching a requested crystal system. Our results also expose the difficulty of optimizing these objectives jointly.

We compare against recent crystal generation models (Chemeleon2, DiffCSP, MatterGen, CrystalFormer, WyFormer, PLaID++, and Crystal-Text-LLM re-trained on the same base model) with LeMat-GenBench (Betala et al., 2025), which uses an interatomic potential and novelty reference that differ from those used in our rewards. We also examine how strongly such comparisons depend on the evaluation procedure, which we examine with two further benchmarks in Section H.9.

2 Background

Related Work Following CDVAE (Xie et al., 2021) and SyMat (Luo et al., 2023), equivariant diffusion models have been a popular paradigm for crystal generation. DiffCSP (Jiao et al., 2023), MatterGen (Zeni et al., 2023), and UniMat (Yang et al., 2023) jointly denoise continuous and discrete crystal variables. Text-conditioned models such as TGDMat (Das et al., 2025), CrysLLMGen (Khasta-

gir et al., 2026), and Chemeleon (Park et al., 2025) enable generation from natural language, while DiffCSP++ (Jiao et al., 2024) and SymmCD (Levy et al., 2025) incorporate space-group symmetry through Wyckoff positions. Flow matching (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023; Liu et al., 2023) has also emerged as an alternative for crystal generation (Miller et al., 2024; Sriram et al., 2024). Transformers trained on Wyckoff-based representations, including CrystalFormer (Cao et al., 2025), WyFormer (Kazeev et al., 2025), and Matra-Genoa (De Breuck et al., 2026), show that discrete space-group-and-Wyckoff sequences provide a strong generative prior. Large language models have also been applied to materials generation (Van et al., 2025): CrystalLLM (Antunes et al., 2024) and Crystal-Text-LLM (Gruver et al., 2024) fine-tune token-based models on CIF-style unit-cell descriptions, while Crystal-ICL (Wang et al., 2025) injects symmetry information through prompts. Reinforcement learning has been used for improved stability and property in crystal generation LLMs (Govindarajan et al., 2025; 2024; Park & Walsh, 2025; Mohanty et al., 2026; Xu et al., 2025). PLaID++ (Xu et al., 2025) fine-tunes an LLM on Wyckoff strings before preference optimization. CrysTune instead explicitly teaches Wyckoff tables through lookup and canonicalization tasks, measures the resulting knowledge, applies on-policy GRPO with a multi-term reward while studying its failure modes, and evaluates all methods under a unified benchmark protocol.

Space Groups and Wyckoff Positions A crystal $\mathcal{C}$ with N atoms in its unit cell is characterized by lattice parameters (l_1, l_2, l_3) and angles (α, β, γ). Atomic species are denoted by $\mathbf{A} = \{a_i\}_{i=1}^{N}$, and positions by fractional coordinates $\mathbf{X} = \{\mathbf{x}_i; \mathbf{x}_i \in [0,1)^3\}_{i=1}^{N}$, where $\mathbf{x}_i = (x_i, y_i, z_i)$. The unit cell repeats along the three lattice vectors, so each atom has periodic images generated by integer cell translations. In three dimensions, crystals are classified into seven crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic, based on their symmetries and shapes. They are further categorized into space groups G, indexed by $K \in \{1, \ldots, 230\}$, which describe the full set of symmetry operations compatible with the periodic lattice. Each element of G can be written as $g = (\mathbf{R}, \mathbf{t})$, where $\mathbf{R}$ is a lattice-preserving linear symmetry operation and $\mathbf{t}$ is a translation vector. It acts on a fractional coordinate $\mathbf{r}$ as $g \cdot \mathbf{r} = \mathbf{R}\mathbf{r} + \mathbf{t} \pmod{\mathbf{1}}$. Operations with $\mathbf{R} = \mathbf{I}$ are pure translations, while nontrivial $\mathbf{R}$ includes rotations, reflections, inversions, rotoinversions, screw rotations, and glide reflections.

Within each space group, there exist symmetry-related "slots" in the unit cell, formally orbits $\mathcal{O}_G(\mathbf{r}) = \{g \cdot \mathbf{r} : g \in G\}$. These slots are grouped by symmetry and labeled as Wyckoff positions. Each label, such as $1a$, $2b$, or $4c$, consists of a multiplicity m, indicating the number of equivalent copies per unit cell, and a letter ℓ identifying the slot type, assigned alphabetically from the most symmetric slot "a" to the least symmetric general position. Higher-symmetry slots produce fewer copies and impose coordinate constraints, whereas lower-symmetry slots allow more free coordinates and generate more copies. Each Wyckoff position has a coordinate template $\tau(K, \ell)$ of free parameters $\{x, y, z\}$ and rational constants, e.g., $\tau(174, 3j) = (x, y, 0)$ and $\tau(167, 12c) = (0, 0, z)$, from which all copies of an atom are generated (Tables 18 and 19). Across the 230 space groups there are 1,731 distinct (space group, Wyckoff position) pairs. Thus, a single representative per orbit suffices to reconstruct the remaining $m - 1$ elements by applying the operations of G. Wyckoff representations are compact and sequential, ordering atoms by decreasing site symmetry.

3 METHODS

CrysTune has three stages (Figure 1): crystals are serialized into a canonical Wyckoff string (Section 3.1); a pretrained LLM is fine-tuned on a curriculum of generation and two auxiliary symmetry tasks (Section 3.2); and the resulting policy is fine-tuned with GRPO (Section 3.4). Generated strings are parsed into structures with symmetry-aware corrections (Section 3.3).

3.1 CANONICAL WYCKOFF STRING REPRESENTATION

Following CrystalFormer (Cao et al., 2025), we describe a crystal by its space group and one representative atom per Wyckoff site, and we serialize this description into a single space-separated string $s = \left[n_G \,\|\, \{w_i, a_i, x_i, y_i, z_i\}_{i=1}^{N_s} \,\|\, l_1, l_2, l_3, \alpha, \beta, \gamma \right]$, where $n_G \in \{1, \ldots, 230\}$ is the space group number, and for each of the N_s symmetry-inequivalent sites w_i is the Wyckoff label (multiplicity and letter, e.g. $4\mathtt{a}$), a_i the element symbol, and (x_i, y_i, z_i) the fractional coordinates of one orbit representative, followed by the lattice lengths and angles. Sites are ordered by Wyckoff letter, i.e.

from the most to the least symmetric position. Because every site is expanded by the space group operations, the string is much shorter than a CIF-style atom list and makes the symmetry of the structure explicit. While CrystalFormer treats the discrete and continuous fields with separate heads of a custom transformer, we write all fields as plain text so that an off-the-shelf LLM tokenizer can be used without new tokens, as in Crystal-Text-LLM (Gruver et al., 2024). To obtain the string from a CIF, we standardize the cell with `spglib`[1] and identify the Wyckoff sites with PyXtal (Fredericks et al., 2021). A single crystal admits many equivalent strings, because every atom may be written at any point of its orbit and the whole cell may be shifted by a lattice-centering translation. We call a string *canonical* when it follows the conventions of the International Tables for Crystallography (Wondratschek et al., 2004): **(i)** the origin is the trivial translation $\mathbf{t} = \mathbf{0}$, e.g. $(0, 0, 0)+$ rather than $(\frac{2}{3}, \frac{1}{3}, \frac{1}{3})+$ for space group 167 (Table 19); and **(ii)** each site uses the first listed orbit representative, e.g. $(0, 0, \frac{1}{4})$ for 6a and $(0, 0, z)$ for 12c. Training targets are always canonical.

3.2 Supervised Fine-Tuning with Symmetry Auxiliary Tasks

We fine-tune Qwen-3.5-9B (Qwen Team, 2026) with LoRA (Hu et al., 2022) on the 27,136 crystals of the MP-20 training set (Jain et al., 2013; Xie et al., 2021), a Materials Project subset with at most 20 atoms per cell. Each example is a single user turn rendered with the model's chat template (thinking disabled) and a response; the loss is the standard causal language-modeling cross-entropy over the response tokens. The prompts for each task are given in Section B.

Task 1: Generation. The prompt describes the output format and, as in Crystal-Text-LLM, a random subset of properties of the target crystal drawn from formation energy, band gap, energy above hull, space group number, crystal system, and density. For each training example we draw the number of included properties k uniformly from $\{0, \ldots, 6\}$ and then a random subset of k properties, so that each property is present in half of the prompts and one prompt in seven is unconditional. The target is the canonical string $s(\mathcal{C})$.

Task 2: Wyckoff lookup. Generating a valid string requires knowing, for every (n_G, ℓ) pair, the coordinate template $\tau(n_G, \ell)$, e.g. $\tau(167, 6a) = (0, 0, \frac{1}{4})$ and $\tau(167, 18e) = (x, 0, \frac{1}{4})$. In plain generation this knowledge is supervised only indirectly through coordinate values, and Section 4.1 shows that pretrained models know little of it. We therefore add a lookup task: given the instruction, a space group number, and a Wyckoff label, the model must emit the three-component template (e.g. `x 0 1/4`). We enumerate all 1,731 (n_G, ℓ) pairs of the three-dimensional space groups (in the `spglib` standard setting) and cycle through a fixed permutation of them, so that rare positions are seen approximately as often as common ones.

Task 3: Canonicalization. The same crystal can be written in many symmetry-equivalent ways: any atom may be written at another point of its orbit, obtained by applying a symmetry operation $g \in G$ of the space group to its coordinates, and the whole cell may be shifted by a lattice-centering translation. For example, the 12c representative $(0, 0, z)$ of space group 167 may equally be written as $(0, 0, -z + \frac{1}{2})$ or, after a centering shift, as $(\frac{2}{3}, \frac{1}{3}, z + \frac{1}{3})$ (Table 19). In this task the model receives the string of a training crystal in which the coordinates of some or all sites have been rewritten in one of these equivalent forms, and must output the canonical string. Most inputs (80%) rewrite each site independently with a random group operation of its own Wyckoff position, and the remaining 20% shift all sites by the same centering translation. The task therefore teaches the canonical conventions of Section 3.1 explicitly: different coordinate strings describe the same crystal and must map to one canonical form. Sampling details are in Section C.

Curriculum and data loading. Generation is the dominant task throughout, rising from 65% of the examples in the first epoch to 80% in the last. The Wyckoff table is small and is learned quickly, so the lookup share starts at 20% and halves each epoch down to a floor of 5%. Canonicalization is harder and stays at 15% in every epoch. These proportions are fixed before each epoch rather than drawn anew at every step, and within an epoch each crystal is used at most once and in a single task. We train for four epochs (3,392 steps at an effective batch size of 32) and use the final checkpoint. Section B.1 gives the per-epoch counts and the assignment of crystals to tasks, Section E the hyperparameters, and Figure 3 the training and validation curves.

[1] A library for finding and applying crystal symmetry operations. (Shinohara et al., 2023)

3.3 Inference and Symmetry-Aware Parsing

At inference time, the fine-tuned model π_θ samples a string $\hat{s} \sim \pi_\theta(\cdot \mid p)$ from an unconditional or property-conditioned prompt p. In a Wyckoff string, the multiplicities, the pinned coordinates, and the expansion of each site into its orbit are all determined by the declared space group and Wyckoff letters, and an LLM that predicts these fields token by token may occasionally contradict them, for instance by emitting `6a` where the table requires `2a` or a coordinate of 0.02 on an axis that the template pins to 0. Rather than discarding every string that breaks a rule, we aim to retain what the model has learned: the parser $\mathcal{P}(\hat{s})$ treats $\hat{s}$ as a proposal and projects it onto the "nearest" structure consistent with the declared space group. Strings that cannot be interpreted at all are rejected and resampled, similar to PLaID++ (Xu et al., 2025). This step is specific to letting an LLM emit a full Wyckoff string; Section H.1 explains why other Wyckoff-based generators do not need it.

Hard validity checks. The parser first rejects strings it cannot interpret: those whose fields do not form a space group, a whole number of five-field site records, and six lattice parameters, or that contain an out-of-range space group, an unparseable number, an unknown element or one with $Z > 94$, or a malformed Wyckoff label (Section H.1). After the corrections below, a string is also rejected if `pymatgen` cannot build a structure from it, which catches cases such as overlapping atoms; these rejections lower the validity we report.

Symmetry-aware corrections. Three corrections use the same Wyckoff tables as training: **C1: multiplicity correction**, $m_i = \mathrm{Mult}(\hat{n}_G, \hat{\ell}_i)$, so a label such as "$6a$" in space group 229 is read as "$2a$"; **C2: fixed-coordinate snapping**, where coordinates pinned by $\tau(\hat{n}_G, \hat{\ell}_i)$ are overwritten with their constants while free axes are unchanged; and **C3: orbit expansion**, where the corrected representative is expanded into all m_i equivalent atoms before writing the CIF (Section J). C2 enforces the canonical representative even when a generated coordinate would match another template of the same orbit, e.g. "`18e Cu 0.0 0.30 0.25`" in space group 167 is snapped to $(x, 0, \frac{1}{4})$ although it matches $(0, x, \frac{1}{4})$; this might create overlaps with other orbits, which the structure check above then rejects. In practice the corrections are rarely triggered: for the 2,500 evaluated CrysTune$^{\text{SFT}}$ samples, 0.2% of sites had an inconsistent multiplicity and 0.5% a drifted pinned coordinate, and 1.1% of sampled strings were rejected; the rates are similar for the variants trained without auxiliary tasks and for the RL policy (Section H.2).

3.4 RL Fine-Tuning with GRPO

The four-epoch SFT model initializes both the policy and the KL reference for GRPO. For unconditional generation, every prompt is the property-free generation instruction; for crystal-system conditioning, the prompt set consists of the seven crystal-system instructions. Each step samples 64 prompts and $K = 5$ rollouts per prompt at temperature 1.0, parses each rollout with $\mathcal{P}$ (rejected strings receive the minimum reward -1), scores the resulting structures, and updates the policy with Equation (3) plus a KL penalty to the SFT reference and an entropy bonus in the actor loss. In preliminary runs with a learning rate of 10^{-6}, KL coefficient 10^{-3}, and no entropy bonus, the policy collapsed onto a few prototypes within 120 steps (Section H.4); the main runs therefore use full-parameter fine-tuning at a learning rate of 6×10^{-7}, KL coefficient 10^{-2}, and entropy coefficient 10^{-3}, and Section H.4 compares LoRA and a learning rate of 3×10^{-7}.

Reward. The reward of a parsed structure $\mathcal{C}$ is a weighted sum of four verifiable terms,

$$R(\mathcal{C}) = \omega_E\, \tilde{r}_E(\mathcal{C}) + \omega_C\, r_C(\mathcal{C}) + \omega_X\, \tilde{r}_X(\mathcal{C}) + \omega_S\, \tilde{r}_S(\mathcal{C}), \tag{1}$$

namely an *energy* term $r_E = -\mathrm{clip}(\Delta E_{\text{hull}}, 0, 1)$ that rewards structures close to the convex hull, a *creativity* term r_C that rewards structures that are unique and novel, and *composition-* and *structure-diversity* terms r_X, r_S that reward rollouts that move the batch distribution towards that of the training set in a learned latent space; $\tilde{r}$ denotes batch min–max normalization and weights are given in Table 7. The design follows the reward of Chemeleon2 (Park & Walsh, 2025), with full definitions in Section F, and departs from it in the energy and creativity terms.

Energy. ΔE_{hull} is computed with MACE-OMAT (Batatia et al., 2025), trained on the OMat24 dataset (Barros-Luque et al., 2026), against a phase diagram built from the LeMat-Bulk MACE-

OMAT hull (Siron et al., 2025) (168k entries), so that the potential and the hull are consistent; the original Chemeleon2 reward instead pairs MACE-MPA with a Materials Project hull.

Creativity. A rollout is *unique*, $u(\mathcal{C}) = 1$, if it matches no other structure (by `StructureMatcher`) in the current batch nor in a first-in-first-out (FIFO) buffer of the 1,000 most recently parsed structures, and *novel*, $v(\mathcal{C}) = 1$, if it matches no structure of the same reduced formula in a reference set. The creativity term is

$$r_C(\mathcal{C}) = \begin{cases} 1, & u(\mathcal{C}) \wedge v(\mathcal{C}), \\ 0, & \neg u(\mathcal{C}) \wedge \neg v(\mathcal{C}), \qquad \kappa = 2, \\ d(\mathcal{C})/(\kappa + d(\mathcal{C})), & \text{otherwise}, \end{cases} \tag{2}$$

where $d(\mathcal{C})$ is the minimum average-minimum-distance (AMD) (Widdowson et al., 2022) between $\mathcal{C}$ and the same-formula structures in the batch, the buffer, and the reference. We make three changes relative to Chemeleon2. The buffer is added so that uniqueness is measured across steps rather than within a batch alone. The novelty reference is the 846k-structure Alexandria+MP set of MatterGen rather than the 45k-structure MP-20 set, which encourages exploration and makes the training signal closer to the novelty measured at evaluation. When a rollout is unique but not novel, or novel but not unique, Chemeleon2 awards the raw distance d, which has no upper bound. We award $d/(\kappa + d)$ instead, which stays below 1 and therefore below the credit given to a structure that is both. The two diversity terms follow Chemeleon2 unchanged, and for crystal-system conditioning we add a binary term for matching the requested system. Section F defines every term.

4 RESULTS

We organize the experiments around three questions.

- **Q1:** *Do pretrained LLMs know the Wyckoff tables, and does fine-tuning on Wyckoff strings with auxiliary symmetry tasks teach them?*
- **Q2:** *Given the same base model, does the Wyckoff representation generate better crystals than CIF, and how does it compare with recent crystal generation models?*
- **Q3:** *Can GRPO improve the SFT policy without mode collapse, and does it improve property-conditioned generation?*

Models. CrysTune$^{\text{SFT}}$ is Qwen-3.5-9B fine-tuned with the full three-task curriculum (Section 3.2). CrysTxtLLM is our re-implementation of Crystal-Text-LLM (Gruver et al., 2024): the same Qwen-3.5-9B model, LoRA configuration, training set, and property prompts, but trained on its CIF-style string of atom types, fractional coordinates, and lattice parameters, so that the two differ only in representation and auxiliary tasks. CrysTune$^{\text{RL}}$ is CrysTune$^{\text{SFT}}$ after 300 GRPO steps with the reward and regularization of Section 3.4. We compare with recent crystal generation models, using the authors' released checkpoints trained on MP-20 and their code to generate 2,500 unconditional samples per model: Chemeleon2 (Park & Walsh, 2025), DiffCSP (Jiao et al., 2023), MatterGen (Zeni et al., 2023), CrystalFormer (Cao et al., 2025), WyFormer (Kazeev et al., 2025), and PLaID++ (Xu et al., 2025). Ablations of the auxiliary tasks and of the RL ingredients are in Sections H.3 and H.4.

Evaluation protocol. All models are evaluated with LeMat-GenBench (Betala et al., 2025) under one configuration with 2,500 generated structures each: novelty and uniqueness by `StructureMatcher` against the 5.3M-structure LeMat-Bulk training split (Siron et al., 2025), and stability from relaxation with ORB-v3 (Rhodes et al., 2025) against the LeMat-Bulk ORB hull ($\Delta E_{\text{hull}} \leq 0$; metastable ≤ 0.1 eV/atom). SUN and MetaSUN are the fractions of valid structures that are stable (or metastable), unique, and novel. Neither the evaluation potential nor the novelty reference is the one used in our reward. LLM samples are drawn at temperature 0.7 and top-$p = 1.0$. Section K gives the metric definitions, sampling details, and parser rejection rates.

4.1 DO LLMS KNOW THE WYCKOFF TABLES? (Q1)

Pretrained Qwen-3.5 models recall only a small fraction of the 1,731 templates (Table 1): the 9B model is correct on 9.7% of cases on average and 17.9% in at least one of five samples, and even the

Table 1: Wyckoff lookup benchmark. Given a space group and Wyckoff label, the model must return the canonical coordinate template (e.g. `167 18e` → `x 0 1/4`); answers are scored by symbolic equivalence over the 1,729 positions not shown as in-context examples. Pretrained models are run in non-thinking mode and sampled five times per case (temperature 0.6). The fine-tuned models are decoded greedily. Thinking-mode results and per-degree-of-freedom accuracies are in Section G.

Model	**Training**	**mean@5 (%)**	**pass@5 (%)**
Qwen-3.5-0.8B	pretrained	0.35	1.74
Qwen-3.5-2B	pretrained	8.87	19.61
Qwen-3.5-4B	pretrained	5.26	11.86
Qwen-3.5-9B	pretrained	9.74	17.87
Qwen-3.5-27B	pretrained	15.88	28.51
Qwen-3.5-9B	CrysTune SFT (greedy; CrysTune$^{\text{SFT}}$)	65.76	

Table 2: LeMat-GenBench evaluation of 2,500 structures per model under one protocol (StructureMatcher, LeMat-Bulk reference, ORB-v3 hull). Uniqueness, novelty, stability, and (Meta)SUN rates use valid structures as the denominator; #SUN is the absolute count. Diversity is element and space-group coverage. Higher is better for every column except the mean energy above hull (↓). Best value per column in **bold**. CrysTxtLLM is Crystal-Text-LLM re-trained on Qwen-3.5-9B.

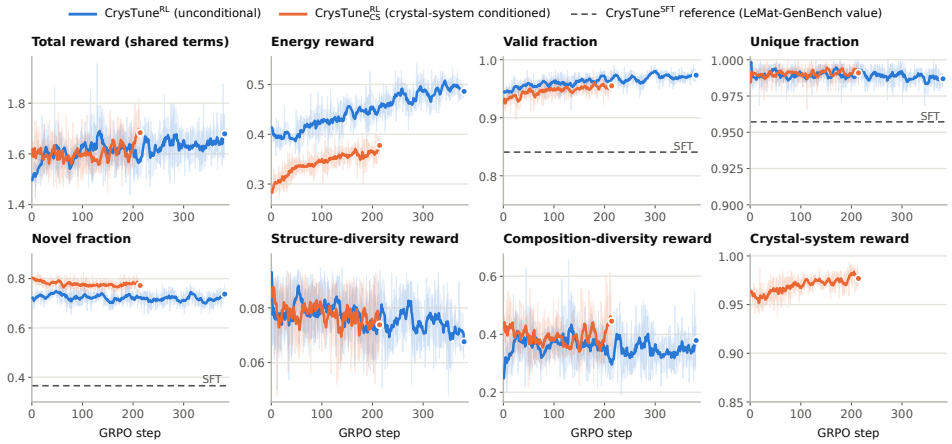
	Generation quality (%)			*Stability (%)*		*Discovery*			*Energy*	*Diversity*	
Model	Valid	Unique	Novel	Stable	Metast.	SUN (%)	#SUN	MetaSUN (%)	$\bar{E}_{\text{hull}}$ ↓	Elem.	SG
Crystal generation models											
Chemeleon2	95.48	91.20	76.54	0.00	37.70	0.00	0	**19.86**	**0.165**	0.653	0.035
DiffCSP	95.48	99.08	68.96	0.59	32.55	0.21	5	8.46	0.272	**0.737**	0.135
MatterGen	**96.08**	99.50	73.44	0.42	36.14	0.12	3	15.03	0.189	0.644	0.161
CrystalFormer	69.32	98.73	43.05	0.58	43.10	0.06	1	3.12	0.690	0.720	0.500
WyFormer	77.72	**99.74**	**79.46**	0.05	4.22	0.00	0	0.15	2.296	0.712	**0.522**
PLaID++	93.80	79.62	30.23	0.90	**56.63**	0.21	5	4.39	0.290	0.712	0.265
LLM fine-tuning, Qwen-3.5-9B											
CrysTxtLLM	79.12	94.59	38.27	0.46	18.86	0.10	2	0.40	0.789	0.703	0.200
CrysTune$^{\text{SFT}}$	84.08	95.72	36.58	1.62	50.48	**0.38**	**8**	2.81	0.488	0.720	0.309
CrysTune$^{\text{RL}}$	90.04	96.05	34.61	**1.69**	48.56	0.36	**8**	3.55	0.399	0.729	0.257

27B model reaches only 28.5% pass@5, with no consistent improvement from 2B to 9B. Enabling the models' thinking mode does not help, since the reasoning rarely terminates within the token budget (Section G). Fine-tuning closes most of the gap: CrysTune$^{\text{SFT}}$ answers 65.8% of held-out cases correctly under greedy decoding, with accuracy rising steadily over the four training epochs (Figure 3). Errors concentrate on templates whose coordinates depend on one another, such as $(x, x+\frac{1}{2}, \frac{1}{4})$, recalled at 35% against 97% for the general position (x, y, z). Whether this knowledge improves generation is examined by the ablation in Section H.3 and summarized below.

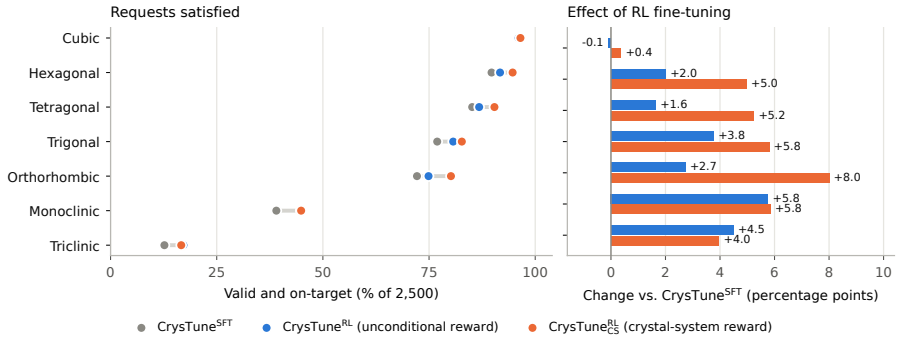
4.2 Comparison with Baselines (Q2)

Table 2 compares all models under one protocol. We draw three conclusions.

Representation matters at a fixed base model. CrysTxtLLM and CrysTune$^{\text{SFT}}$ share the base model, adapter, data, and prompts and differ in representation and auxiliary tasks. CrysTune$^{\text{SFT}}$ is better on validity (84.1 vs. 79.1%), stability (1.62 vs. 0.46%), SUN (0.38 vs. 0.10%, 8 vs. 2 structures), MetaSUN (2.81 vs. 0.40%), and space-group coverage (0.31 vs. 0.20), and marginally lower on novelty (36.6 vs. 38.3%). The ablation in Section H.3 indicates that most of this gain comes from the representation itself: a model trained on generation alone reaches the same SUN count (8) with higher validity and stability but lower novelty, and adding the two auxiliary tasks raises novelty and site-number diversity at a small cost in stability, without a clear net gain under these metrics. The auxiliary tasks make the model a reliable Wyckoff oracle (Section 4.1) and halve the number of invalid Wyckoff letters in generated strings (Section H.2), but they do not by themselves raise SUN.



(a) GRPO training curves, stitched across resumed Slurm segments. Faint lines are per-step batch means over 320 rollouts, bold lines a trailing 10-step mean with a ± 1 standard-deviation band. The total reward excludes the conditioned run's crystal-system term so that the two runs share a scale. Valid, unique, and novel fractions are the reward pipeline's own measurements on parsed rollouts; the dashed CrysTune^{SFT} lines are the LeMat-GenBench values of Table 2, which use stricter validity checks and a larger novelty reference, so they are a reference point rather than the same quantity.



(b) Crystal-system-conditioned generation: for each requested crystal system, the fraction of 2,500 generations that are valid and match the request. Per-system counts are in Table 17.

Figure 2: GRPO fine-tuning of CrysTune^{RL} (unconditional) and CrysTune$^{RL}_{CS}$ (crystal-system conditioned): training behaviour (a) and conditional generation of the resulting policies (b).

Compared with recent crystal generation models, CrysTune improves stability at the cost of novelty. CrysTune^{SFT} has the highest stable rate (1.62% vs. at most 0.90%), SUN rate (0.38% vs. 0.21% for DiffCSP and PLaID++), and SUN count among all models, although at 2,500 samples per model these two rates rest on relatively few structures, unlike the metastable rate (50.5%), which is second only to PLaID++. The stable structures also share a marked preference for rare-earth elements, which we examine in Section H.8. Its novelty is below that of the diffusion models and comparable to the two other Wyckoff-based LLMs, PLaID++ and CrysTxtLLM, and its MetaSUN rate is well below those of Chemeleon2 and MatterGen. Diffusion models thus explore more broadly and produce many metastable novel structures, whereas CrysTune produces fewer but more often on the hull. In space-group coverage CrysTune sits between the CIF-based CrysTxtLLM and the Wyckoff-native transformers, which generate a space group before any coordinate. The ranking also depends on the reference set: with the MP-20 training set as the novelty reference, as used in several prior papers, the novelty of CrysTune^{SFT} is 74% and that of CrysTune^{RL} 77%, and their SUN rates roughly double (Section H.6), which changes the ordering of several methods; the choice of hull and stability threshold has an even larger effect (Section H.9).

4.3 RL FINE-TUNING (Q3)

Unconditional generation. CrysTune^{RL} uses the reward and regularization of Section 3.4. After 300 steps it improves validity (84.1 to 90.0%), stability (1.62 to 1.69%), and MetaSUN (2.81 to 3.55%, 59 to 80 structures) over CrysTune^{SFT}, while uniqueness (96.1%), element diversity, and the SUN count (8) are unchanged, and novelty and space-group coverage drop marginally (36.6 to 34.6% and 0.31 to 0.26). Figure 2a shows the same picture during training: the energy reward rises steadily while the unique and novel fractions stay flat. Comparing space-group distributions (Figure 4), CrysTune^{RL} retains most of the SFT model's coverage (66 versus 86 of the 230 space groups) and sharpens its preference for the cubic space group 225 from 23% to 30% of samples. These gains are modest in absolute terms, and obtaining them at all depends on the exploration measures of Section 3.4. Removing those measures makes the policy optimize the reward far more aggressively: it reaches 99% validity and a much lower mean energy above hull, but concentrates on a handful of prototypes, so that uniqueness falls to 67%, novelty to 15%, space-group coverage to 0.017. Section H.4 adds the measures back one at a time at matched training steps: bounding the creativity term and strengthening the KL and entropy regularization recover uniqueness but not novelty or coverage, the cross-batch uniqueness buffer recovers both, and the better-matched energy model and larger novelty reference then restore the validity gain. LoRA and full fine-tuning behave similarly, whereas a higher learning rate trades novelty for validity.

Crystal-system-conditioned generation. Prompting CrysTune^{SFT}, CrysTune^{RL}, and the policy trained with the additional crystal-system reward (CrysTune$^{RL}{}_{CS}$, Section 3.4) with each of the seven crystal systems, the share of generations that are valid and match the request rises from 67.4% to 70.3% and 72.3% respectively (Figure 2b). The margin available to RL is narrow: half of the SFT generation prompts already carry the crystal system, so the supervised model follows the instruction well and the crystal-system reward begins near 0.96 (Figure 2a). All three models satisfy high-symmetry requests far more often than low-symmetry ones. Against the unconditional policy, the crystal-system reward helps most on mid-symmetry targets, by 3 to 5 percentage points for hexagonal, tetragonal, and orthorhombic requests, and leaves monoclinic and triclinic essentially unchanged. Conditioning on a property that the supervised model already handles therefore leaves limited room for improvement, and evaluating properties that are harder to satisfy from the prompt alone would be a more informative test of RL fine-tuning.

5 DISCUSSION AND CONCLUSION

We set out to investigate whether a richer, symmetry-aware representation, paired with auxiliary tasks that teach the underlying crystallographic rules, helps an LLM generate better crystals. The supervised model does learn the Wyckoff tables, improving from below 10% to 66% exact-match recall, and at a fixed base model and training set the Wyckoff representation produces more stable and more symmetry-diverse crystals than a CIF-style one. Reinforcement learning then improves validity and the yield of metastable structures that are unique and novel, but only under exploration measures that prevent the policy from concentrating on a few prototypes, and the gains on any one property are accompanied by little or no movement on the others. Improving several properties at once therefore remains difficult, and we regard this work as a step towards symmetry-aware training and RL fine-tuning of LLMs for crystal generation rather than an attempt to outperform established generative models. We also note that such comparisons are sensitive to the evaluation procedure: the choice of convex hull, interatomic potential, novelty reference, and stability threshold changes both the absolute numbers and the ordering of methods, as we show by re-evaluating the same structures under three protocols (Sections H.6 and H.9). Several directions follow from this study: validating a subset of generated candidates with density functional theory, training on datasets larger than MP-20 and assessing how performance scales with model size, and exploring more aggressive exploration strategies that raise novelty and coverage without eroding the validity and stability the supervised model already provides.

AI USE STATEMENT

Generative AI tools assisted in preparing this manuscript: refining the writing, surveying candidate related work, and drafting portions of the text. Coding assistants also helped implement parts of

the training and evaluation pipelines and produce figures from computed results. Every AI-drafted passage and all such code were reviewed by the authors, and every number reported here comes from the authors' own experimental runs. The authors accept full responsibility for the content of this paper.

REFERENCES

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *ICLR*, 2023.

Luis M Antunes, Keith T Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling. *Nature Communications*, 15(1):10570, 2024.

Luis Barros-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, Matt Uyttendaele, C Lawrence Zitnick, and Zachary W Ulissi. The open materials 2024 (omat24) inorganic materials dataset and models. *Nature Computational Science*, pp. 1–11, 2026.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

Siddharth Betala, Samuel P Gleason, Ali Ramlaoui, Andy Xu, Georgia Channing, Daniel Levy, Clémentine Fourrier, Nikita Kazeev, Chaitanya K Joshi, Sékou-Oumar Kaba, et al. Lematgenbench: A unified evaluation framework for crystal generative models. *arXiv preprint arXiv:2512.04562*, 2025.

Zhendong Cao, Xiaoshan Luo, Jian Lv, and Lei Wang. Space group informed transformer for crystalline materials generation. *Science Bulletin*, 2025.

Kishalay Das, Subhojyoti Khastagir, Pawan Goyal, Seung-Cheol Lee, Satadeep Bhattacharjee, and Niloy Ganguly. Periodic materials generation using text-guided joint diffusion model. In *ICLR*, 2025.

Pierre-Paul De Breuck, Hashim A. Piracha, Gian-Marco Rignanese, and Miguel A. L. Marques. A generative material transformer using wyckoff representation. *npj Computational Materials*, 12(1): 60, Jan 2026. ISSN 2057-3960. doi: 10.1038/s41524-025-01940-8. URL https://doi.org/10.1038/s41524-025-01940-8.

Scott Fredericks, Kevin Parrish, Dean Sayre, and Qiang Zhu. Pyxtal: A python library for crystal structure generation and symmetry analysis. *Computer Physics Communications*, 261:107810, 2021.

Prashant Govindarajan, Santiago Miret, Jarrid Rector-Brooks, Mariano Phielipp, Janarthanan Rajendran, and Sarath Chandar. Learning conditional policies for crystal design using offline reinforcement learning. *Digital Discovery*, 3:769–785, 2024. doi: 10.1039/D4DD00024B. URL http://dx.doi.org/10.1039/D4DD00024B.

Prashant Govindarajan, Mathieu Reymond, Antoine Clavaud, Mariano Phielipp, Santiago Miret, and Sarath Chandar. Crystalgym: A new benchmark for materials discovery using reinforcement learning. *arXiv preprint arXiv:2509.23156*, 2025.

Nate Gruver, Anuroop Sriram, Andrea Madotto, Andrew Gordon Wilson, C. Lawrence Zitnick, and Zachary Ward Ulissi. Fine-tuned language models generate stable inorganic materials as text. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=vN9fpfqoP1.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *arXiv preprint arXiv:2309.04475*, 2023.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. *arXiv preprint arXiv:2402.03992*, 2024.

Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff transformer: Generation of symmetric crystals. *arXiv preprint arXiv:2503.02407*, 2025.

Subhojyoti Khastagir, Kishalay Das, Pawan Goyal, Seung-Cheol Lee, Satadeep Bhattacharjee, and Niloy Ganguly. LLM Meets Diffusion: A Hybrid Framework for Crystal Material Generation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=E6gwPtWjb1.

Woosuk Kwon. *vLLM: An Efficient Inference Engine for Large Language Models*. PhD thesis, UC Berkeley, 2025.

Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. SymmCD: Symmetry-preserving crystal generation with diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xnssGv9rpW.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. *ICLR*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *ICLR*, 2023.

Youzhi Luo, Chengkai Liu, and Shuiwang Ji. Towards symmetry-aware generation of periodic materials. *Advances in Neural Information Processing Systems*, 2023.

Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. *arXiv preprint arXiv:2406.04713*, 2024.

Trupti Mohanty, Maitrey Mehta, Hasan M Sayeed, Bat-El Oded, Itay Pitussi, Arie Borenstein, Vivek Srikumar, and Taylor D Sparks. Crystext: a generative ai approach for text-conditioned crystal structure generation using llm. *Integrating Materials and Manufacturing Innovation*, pp. 1–15, 2026.

Hyunsoo Park and Aron Walsh. Guiding generative models to uncover diverse and novel crystals via reinforcement learning. *arXiv preprint arXiv:2511.07158*, 2025.

Hyunsoo Park, Anthony Onwuli, and Aron Walsh. Exploration of crystal chemical space using text-guided generative artificial intelligence. *Nature Communications*, 16(1):4379, 2025.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024.

Kohei Shinohara, Atsushi Togo, and Isao Tanaka. Algorithms for magnetic symmetry operation search and identification of magnetic space group from magnetic crystal structure. *Acta Cryst. A*, 79(5):390–398, Sep 2023. doi: 10.1107/S2053273323005016. URL https://doi.org/10.1107/S2053273323005016.

Martin Siron, Inel Djafar, Ali Ramlaoui, Etienne du Fayette, Amandine Rossello, Edvin Fako, Matthew McDermott, Felix Therrien, Luis Barroso-Luque, Flaviu Cipcigan, et al. Lematbulk: aggregating, and de-duplicating quantum chemistry materials databases. *arXiv preprint arXiv:2511.05178*, 2025.

Anuroop Sriram, Benjamin Miller, Ricky TQ Chen, and Brandon Wood. Flowllm: Flow matching for material generation with large language models as base distributions. *NeurIPS*, 37:46025–46046, 2024.

Minh-Hao Van, Prateek Verma, Chen Zhao, and Xintao Wu. A survey of ai for materials science: foundation models, llm agents, datasets, and tools. *arXiv preprint arXiv:2506.20743*, 2025.

Ruobing Wang, Qiaoyu Tan, Yili Wang, Ying Wang, and Xin Wang. Crystalicl: Enabling in-context learning for crystal generation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 18440–18455, 2025.

Daniel Widdowson, Marco Mosca, Angeles Pulido, Vitaliy Kurlin, and Andrew I Cooper. Average minimum distances of periodic point sets are fundamental invariants for mapping all periodic crystals. *MATCH Communications in Mathematical and in Computer Chemistry*, 87(3):529–559, 2022.

Hans Wondratschek, Ulrich Müller, and Union internationale de cristallographie. *International tables for crystallography*, volume 1. Wiley Online Library, 2004.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197*, 2021.

Andy Xu, Rohan Desai, Larry Wang, Gabriel Hope, and Ethan Ritz. Plaid++: A preference aligned language model for targeted inorganic materials design. *arXiv preprint arXiv:2509.07150*, 2025.

Mengjiao Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235*, 2023.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya, Jonathan Crabbé, Lixin Sun, Jake Smith, et al. Mattergen: a generative model for inorganic materials design. *arXiv preprint arXiv:2312.03687*, 2023.

A Appendix

B Prompts

All three SFT tasks use a single user turn rendered with the Qwen-3.5 chat template with thinking disabled; the response is the target string followed by the end-of-turn token. The prompts below are reproduced verbatim from the training code.

Generation. The prompt starts with `Below is a description of a bulk material.` followed by zero to six property sentences drawn at random from `The formation energy per atom is {v}.`, `The band gap is {v}.`, `The energy above the convex hull is {v}.`, `The spacegroup number is {v}.`, `The crystal system is {v}.`, and `The density is {v}.`, and then the format instruction:

```
Generate the crystal structure in Wyckoff representation
format.  The output is a single line with the following
fields in order:  (1) space group number (an integer from
1 to 230); (2) for each symmetry-inequivalent site, the
Wyckoff symbol (multiplicity + letter, e.g.  4a or 12e),
the element symbol, and the fractional coordinates x y z.
Fractional coordinates are understood modulo 1 and must be
reported in the canonical interval [0, 1); (3) the lattice
parameters a b c (Angstrom) and alpha beta gamma (degrees).
```

```
List each symmetry-inequivalent atomic orbit exactly once;
do not list duplicate sites or multiple symmetry-equivalent
representatives of the same orbit.  Repeated element symbols
and repeated Wyckoff symbols are allowed when they describe
distinct orbits.  After applying the space-group symmetry
operations and periodic boundary conditions, atoms from
distinct sites must not occupy the same position.  Avoid
unphysically short interatomic distances.
```

Example target: `225 4a Mn 0.0000 0.0000 0.0000 4b Li 0.5000 0.5000 0.5000 8c Ir 0.2500 0.2500 0.2500 5.8894 5.8894 5.8894 90.00 90.00 90.00.`

Wyckoff lookup.

```
Given a space group number and an exact
multiplicity-and-letter Wyckoff symbol, use the Wyckoff
table in the standard setting and origin convention used by
the IUCr CIF symmetry dictionary.  Return the canonical
generator, meaning the first listed coordinate triplet
for that position.  Keep free variables symbolic and fixed
values as exact fractions.
Return exactly three space-separated coordinate expressions.
Do not use commas or add an explanation.
Examples:
Input:  62 4c
Output:  x 1/4 z
Input:  194 2c
Output:  1/3 2/3 1/4
Input:
167 18e
```

Target: `x 0 1/4.`

Canonicalization.

```
The input is a crystal in Wyckoff representation format.
One or more sites use symmetry-equivalent members of their
Wyckoff orbits instead of the canonical representatives.
Convert every site to the canonical first representative
in the same standard space-group setting and origin
convention used by the IUCr CIF symmetry dictionary.
Preserve the space group, Wyckoff symbols, elements, and
lattice parameters.  Output only the canonical crystal
representation.
Input:
225 4a Mn 0.5000 0.0000 0.5000 4b Li 0.5000 0.0000 0.0000
8c Ir 0.2500 0.7500 0.2500 5.8894 5.8894 5.8894 90.00 90.00
90.00
```

Target: `225 4a Mn 0.0000 0.0000 0.0000 4b Li 0.5000 0.5000 0.5000 8c Ir 0.2500 0.2500 0.2500 5.8894 5.8894 5.8894 90.00 90.00 90.00.`

B.1 Curriculum and data loading

An epoch consists of 27,136 examples, one per training crystal, split among the three tasks with proportions $(p_{\text{gen}}, p_{\text{can}}, p_{\text{look}})$ set before the epoch begins. The lookup share follows $p_{\text{look}}(e) = \max(0.05,\, 0.20 \cdot 0.5^{e})$ and $p_{\text{can}} = 0.15$ is constant, giving the counts below; integer counts are obtained by the largest-remainder method. The lookup examples cover the 1,731 positions about three times in the first epoch and about six times over the four epochs.

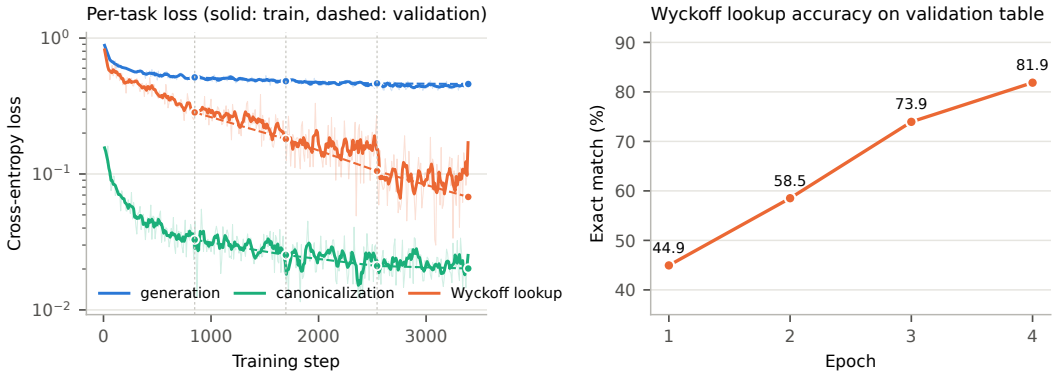


Figure 3: Supervised fine-tuning of Qwen-3.5-9B on the four-epoch curriculum. Left: per-task cross-entropy, with faint lines the logged training values, bold lines a five-point running mean, and dashed lines with markers the end-of-epoch validation values. Dotted vertical lines mark epoch boundaries. Right: exact-match accuracy on the validation copy of the Wyckoff table over all 1,731 positions. This is measured with teacher forcing, meaning the model is scored one position at a time on the ground-truth prefix, so errors cannot compound; it is therefore higher than the free-generation accuracy of Table 1 and should be read as a training diagnostic rather than as a comparable number. The canonicalization loss is low from the first epoch, whereas the lookup loss falls steadily as the model memorizes the table.

Table 3: Task plan of the four-epoch curriculum.

Epoch	p_{gen}	p_{can}	p_{look}	Generation	Canonicalization	Lookup
1	0.65	0.15	0.20	17,639	4,070	5,427
2	0.75	0.15	0.10	20,352	4,070	2,714
3	0.80	0.15	0.05	21,709	4,070	1,357
4	0.80	0.15	0.05	21,709	4,070	1,357

Crystals are assigned to tasks from a fixed random permutation of the 27,136 training crystals: in each epoch the next 4,070 crystals of this permutation are used for canonicalization, so the four epochs canonicalize disjoint sets, and the generation examples of that epoch are drawn from the crystals not used for canonicalization. Each crystal therefore appears at most once per epoch and in only one task. Lookup cases are drawn as a contiguous cyclic slice of a fixed permutation of the 1,731 positions, so every position is seen once before any is repeated. The plan is a deterministic function of the seed and epoch index, which makes a pre-empted run reproducible on resumption.

C Canonicalization Task: Sampling Details

For site i with Wyckoff position ℓ_i, let $W_1^{(\ell_i)}, \dots, W_{m_i}^{(\ell_i)}$ be the affine maps that generate its orbit (Section J), with $W_1^{(\ell_i)}$ the canonical representative. Each canonicalization example is built from one training crystal by one of two perturbations. **Per-site** (probability 0.8): independently for every site, $\mathbf{x}_i$ is replaced by $W_j^{(\ell_i)}\mathbf{x}_i$ for a j drawn uniformly among the maps that move the point, i.e. $W_j^{(\ell_i)}\mathbf{x}_i \neq \mathbf{x}_i$; sites whose orbit has a single distinct point (e.g. multiplicity-one positions) are left unchanged. **Global** (probability 0.2): one non-trivial pure translation $\mathbf{v} \in \mathcal{T}(G) = \{\mathbf{v} \neq \mathbf{0} : (\mathbf{I}, \mathbf{v}) \in G\}$, i.e. a lattice-centering vector, is added to every site, $\mathbf{x}_i \leftarrow (\mathbf{x}_i + \mathbf{v}) \bmod \mathbf{1}$; for space group 167 this set is $\{(\frac{2}{3}, \frac{1}{3}, \frac{1}{3}), (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})\}$. If the chosen perturbation leaves the string unchanged, which happens for the global mode on primitive lattices ($\mathcal{T}(G) = \varnothing$), the other mode is used; if both leave it unchanged, another training crystal is drawn (up to eight attempts). The target is always the canonical string $s(\mathcal{C})$, and the loss is the standard cross-entropy over all response tokens. In a sample of 60 MP-20 crystals, 35 admitted a non-trivial centering translation.

D GROUP RELATIVE POLICY OPTIMIZATION

We fine-tune a pretrained LLM, denoted by the policy π_θ with parameters θ, first with supervised next-token prediction on prompt–response pairs and then with Group Relative Policy Optimization (GRPO) (Shao et al., 2024), a policy-gradient method that needs no value-function critic. For each prompt, GRPO samples a group of K responses $S_1, \ldots, S_K \sim \pi_\theta$ and scores each with a reward R. The rewards are normalized within the group to give the advantage of response i, which is optimized with a PPO-style clipped surrogate objective,

$$A_i = \frac{R(S_i) - \frac{1}{K}\sum_{j=1}^{K} R(S_j)}{\sigma_R}, \qquad J(\theta) = \mathbb{E}\left[\sum_{i=1}^{K} \min\left(r_i(\theta)A_i, \mathrm{clip}(r_i(\theta), 1-\epsilon, 1+\epsilon)A_i\right)\right], \tag{3}$$

where σ_R is the standard deviation of the rewards within the group, $r_i(\theta)$ is the probability ratio between the current policy and the policy that produced the rollout, and clipping to $[1-\epsilon, 1+\epsilon]$ keeps the update within a trust region. Our runs add a KL penalty to the SFT reference policy and an entropy bonus to this objective, with the coefficients of Table 5.

E HYPERPARAMETERS

E.1 SFT HYPERPARAMETERS

Table 4: SFT hyperparameters for CrysTune$^{\text{SFT}}$ (Qwen-3.5-9B, four-epoch curriculum). CrysTxtLLM uses the same values with the CIF-style representation and no auxiliary tasks.

Hyperparameter	**Value**
Base model	`Qwen/Qwen3.5-9B` (bf16, thinking disabled)
Training / validation data	MP-20 train (27,136) / val (9,047)
Epochs / optimizer steps	4 / 3,392 (848 per epoch)
GPUs	4×H100 (DDP)
Per-device batch × grad. accumulation	1×8 (effective 32)
Learning rate / schedule / warmup	1×10^{-4} / cosine / 170 steps
Weight decay / gradient clipping	0.0 / 1.0
Maximum sequence length	2,048 tokens
LoRA rank / alpha / dropout	8 / 32 / 0.05
LoRA target modules	attention, linear-attention, and MLP projections (21.6M trainable)
Task mix (gen / canon / lookup), epoch 1	0.65 / 0.15 / 0.20
Lookup decay / floor	$\times 0.5$ per epoch / 0.05
Canonicalization mode	mixed (per-site 0.8, global 0.2)
Fixed-coordinate loss masking	Disabled
Property conditioning	0–6 properties per prompt, uniform
Seed	42

E.2 GRPO HYPERPARAMETERS

We perform GRPO fine-tuning using VERL (Sheng et al., 2024) on $4\times$H100 GPUs with FSDP2 sharding and bf16 weights; vLLM (Kwon, 2025) handles on-policy rollout generation with tensor parallelism 2. The interatomic potential, phase diagram, novelty index, replay buffer, and VAE featurizer are instantiated once per worker and reused across steps; a batch-level timeout of $180\,\mathrm{s}$ on the energy evaluation maps the batch to the minimum energy reward. The replay buffer is saved with each checkpoint and restored on resume. Table 5 lists the settings of the main runs; Table 6 lists what differs between the unregularized and regularized recipes.

Table 5: GRPO hyperparameters shared by the unconditional and crystal-system-conditional runs of CrysTune$^{\text{RL}}$.

Setting	Value
Policy and KL reference	CrysTune$^{\text{SFT}}$ (four-epoch checkpoint)
Prompts per step / rollouts per prompt K	64 / 5 (320 structures per step)
PPO mini-batch / micro-batch / epochs	16 prompts / 1 / 1
Max prompt / response tokens	512 / 500
Sampling temperature / top-p	1.0 / 1.0
Actor learning rate	6×10^{-7} (main), 3×10^{-7} (ablation), AdamW
Fine-tuning	full parameters (main); LoRA $r = 8$, $\alpha = 32$ (ablation)
KL coefficient / placement	10^{-2} / actor loss (low-variance estimator)
Entropy coefficient	10^{-3}
Steps	320 (20 epochs of 1,024 prompts); extended to 640
Energy model / hull	MACE-OMAT-0 medium / LeMat-Bulk MACE-OMAT hull (168,458 entries)
Novelty reference	Alexandria+MP (845,997 structures, reduced-formula index)
Replay buffer	FIFO, 1,000 parsed structures
Creativity partial credit	$d/(2+d)$
Checkpoint interval	20 steps

Table 6: Differences between the unregularized recipe (CrysTune$^{\text{RL-naive}}$) and the regularized recipe (CrysTune$^{\text{RL}}$).

Ingredient	CrysTune$^{\text{RL-naive}}$	CrysTune$^{\text{RL}}$
Energy model / hull	MACE-MPA-0 / Materials Project	MACE-OMAT-0 / LeMat-Bulk MACE-OMAT
Novelty reference	MP-20 (45k structures)	Alexandria+MP (846k structures)
Uniqueness	within batch only	within batch and 1,000-structure replay buffer
Creativity partial credit	raw AMD distance (unbounded)	$d/(2+d)$ (bounded)
Learning rate	1×10^{-6}	6×10^{-7}
KL coefficient / entropy bonus	10^{-3} / 0	10^{-2} / 10^{-3}

F Reward Details

Table 7: Reward components active during GRPO fine-tuning. The † term is only used in the crystal-system conditional run.

Component	Weight w_c	Normalization	Range (pre-norm)
r_E (energy above hull)	1.0	batch min–max	$[-1, 0]$
r_C (creativity)	1.0	–	$[0, 1]$
r_X (composition MMD)	1.0	batch min–max	$\mathbb{R}$
r_S (structure MMD)	0.1	batch min–max	$\mathbb{R}$
r_{CS} (crystal system) †	1.0	–	$\{0, 1\}$

The total reward is the weighted sum of the components with no global normalization; parse failures and empty structures receive $R_{\min} = -1$ and are not inserted into the replay buffer. A component that raises an exception contributes zero for that structure.

Energy above hull. For each generated structure, the energy is predicted with MACE-OMAT and wrapped as a `PDEntry`; `PatchedPhaseDiagram.get_e_above_hull` against the LeMat-

Bulk MACE-OMAT hull gives ΔE_{hull} in eV/atom. The raw reward is

$$r_E(\mathcal{C}) = -\text{clip}(\max(\Delta E_{\text{hull}}(\mathcal{C}), 0),\ 0,\ 1)\,, \tag{4}$$

so structures on or below the hull receive 0, structures up to 1 eV/atom above it receive $-\Delta E_{\text{hull}}$, and structures above 1 eV/atom or with failed energy evaluations receive -1.

Creativity. A structure is *unique*, $u(\mathcal{C}) = 1$, if it is the first member of its `StructureMatcher` group within the rollout batch and matches no structure of the same reduced formula among the 1,000 most recent parsed structures in the replay buffer; it is *novel*, $v(\mathcal{C}) = 1$, if it matches no structure of the same reduced formula in the Alexandria+MP reference. With $d(\mathcal{C})$ the minimum average-minimum-distance (AMD) (Widdowson et al., 2022) between $\mathcal{C}$ and the same-formula structures of the batch, the buffer, and the reference,

$$r_C(\mathcal{C}) \;=\; \begin{cases} 1, & u(\mathcal{C}) \wedge v(\mathcal{C}), \\ 0, & \neg u(\mathcal{C}) \wedge \neg v(\mathcal{C}), \\ \dfrac{d(\mathcal{C})}{\kappa + d(\mathcal{C})}, & \text{otherwise}, \end{cases} \qquad \kappa = 2. \tag{5}$$

The bounded form guarantees that a structure which is both unique and novel receives at least as much creativity reward as any structure that is only one of the two. With the raw AMD of the original formulation, the partial-credit branch exceeded 1 in 96% of logged steps of our preliminary runs.

Distributional diversity. Let z denote either the structure-level or composition-level Chemeleon2 VAE embedding. With a cubic polynomial kernel $k(x, y) = (x^\top y/d + 1)^3$ we form a biased MMD^2 estimator between the batch and MP-20 reference features and assign each rollout the change in MMD induced by leaving it out of the batch:

$$\widehat{\text{MMD}}^2(\mathbf{z}, \mathbf{z}^{\text{ref}}) \;=\; \tfrac{1}{M(M-1)}\sum_{i\neq j} k(z_i, z_j) + \tfrac{1}{N(N-1)}\sum_{i\neq j} k(z_i^{\text{ref}}, z_j^{\text{ref}}) - \tfrac{2}{MN}\sum_{i,j} k(z_i, z_j^{\text{ref}}), \tag{6}$$

$$r_{\{S,X\}}(\mathcal{C}^{(k)}) \;=\; \widehat{\text{MMD}}^2_{\setminus k}(\mathbf{z}, \mathbf{z}^{\text{ref}}) \;-\; \widehat{\text{MMD}}^2(\mathbf{z}, \mathbf{z}^{\text{ref}}). \tag{7}$$

Positive r_S, r_X indicate rollouts whose removal would increase the MMD to the reference distribution.

Crystal system target matching. For the crystal-system conditional run, each prompt p carries a target crystal system $y_p \in \{\text{cubic}, \ldots, \text{triclinic}\}$. We compute the generated structure's crystal system with `SpacegroupAnalyzer` using `symprec`$= 0.1$ and `angle_tol`$= 5^\circ$, and assign $r_{CS}(\mathcal{C}; y_p) = \mathbf{1}[\text{CS}(\mathcal{C}) = y_p]$.

G Wyckoff Lookup Benchmark

The benchmark enumerates all 1,731 (space group, Wyckoff position) pairs of the three-dimensional space groups with PyXtal (Fredericks et al., 2021) in the `spglib` standard setting and asks for the first listed coordinate triplet. The two in-context examples of the prompt (`62 4c` and `194 2c`) are excluded from scoring, leaving 1,729 cases. Answers are scored by symbolic equivalence (SymPy), so `0.25` and `1/4` are equivalent, and additionally modulo integer translations; format-compliance and truncation rates are also recorded, with errors counted as incorrect. Pretrained models were run with vLLM at temperature 0.6, top-p 0.95, top-k 20, five samples per case, and a 2,048-token limit, in both non-thinking and thinking mode. The fine-tuned model was decoded greedily with a 128-token limit. The pretrained accuracies are recall of PyXtal's selected representative and would be somewhat higher if equivalent representatives in other settings were accepted.

Table 8: Wyckoff lookup accuracy of pretrained Qwen-3.5 models over 1,729 held-out positions (five samples per case). Thinking mode fails because reasoning almost never terminates within the token budget.

Model	Mode	Symbolic mean@5	Symbolic pass@5	Mod-1 pass@5	Answer format	Truncated
0.8B	non-thinking	0.35%	1.74%	2.54%	73.92%	0.00%
2B	non-thinking	8.87%	19.61%	19.84%	77.70%	0.00%
4B	non-thinking	5.26%	11.86%	12.09%	89.62%	3.47%
9B	non-thinking	9.74%	17.87%	18.33%	97.35%	0.03%
27B	non-thinking	15.88%	28.51%	29.27%	100.00%	0.00%
0.8B	thinking	0.01%	0.06%	0.64%	0.00%	100.00%
2B	thinking	0.02%	0.12%	0.12%	0.00%	100.00%
4B	thinking	0.07%	0.35%	0.40%	0.00%	100.00%
9B	thinking	0.05%	0.23%	0.29%	0.00%	100.00%
27B	thinking	0.28%	0.98%	0.98%	0.00%	99.75%

Table 9: Wyckoff lookup accuracy after CrysTune SFT (Qwen-3.5-9B, greedy decoding, 1,729 held-out positions) for the two-epoch and four-epoch checkpoints, and the teacher-forced exact-match accuracy on the validation copy of the full table during the four-epoch run. Degrees of freedom (DOF) is the number of free coordinates in the template.

Quantity	2 epochs	4 epochs (CrysTune$^{\text{SFT}}$)
Symbolic accuracy (1,729 held-out)	61.02%	65.76%
Modulo-one accuracy	61.19%	65.88%
Answer-format compliance / truncation	99.83% / 0.00%	99.94% / 0.00%
Accuracy by template type, CrysTune$^{\text{SFT}}$: all coordinates fixed, e.g. $(0, 0, \frac{1}{4})$	65.7% (404/615)	
independent free variables, e.g. $(x, 0, \frac{1}{4})$	59.9% (480/801)	
coordinates dependent on one another, e.g. $(x, x+\frac{1}{2}, \frac{1}{4})$	34.9% (29/83)	
general position (x, y, z)	97.4% (224/230)	
Mean completion tokens	10.4	10.4
Teacher-forced validation exact match, epochs 1 / 2 / 3 / 4	44.95% / 58.52% / 73.95% / 81.86%	

The teacher-forced metric checks greedy token predictions under the ground-truth prefix over all 1,731 positions and is not interchangeable with the free-generation benchmark, since a model that would derail partway through a free continuation is still scored correct at every later position; the fine-tuned models are decoded greedily whereas the pretrained models are sampled, so the comparison against mean@5 slightly favors the fine-tuned models while the comparison against pass@5 favors the pretrained ones; the gap (65.8% greedy versus 28.5% pass@5 for the best pretrained model) exceeds either bias.

H Additional Results

H.1 Parser hard validity checks

Let $\hat{s} = \left[\, \hat{n}_G \,\|\, \{\hat{w}_i, \hat{a}_i, \hat{x}_i, \hat{y}_i, \hat{z}_i\}_{i=1}^{\hat{N}_s} \,\|\, \hat{l}_1, \hat{l}_2, \hat{l}_3, \hat{\alpha}, \hat{\beta}, \hat{\gamma} \,\right]$ be the generated string split at whitespace into fields. The parser rejects it if it has fewer than eight fields (one site plus the space group and lattice); if the site block cannot be partitioned into complete five-field site records; if $\hat{n}_G \notin \{1, \ldots, 230\}$; if any coordinate or lattice parameter is not a real number; if an element symbol is unrecognized or has $Z > 94$; or if a Wyckoff label does not match "`<integer><letter>`". The letter set is $\{\mathrm{a}, \ldots, \mathrm{z}, \mathrm{A}\}$ because space group 47 has 27 positions, the last of which is labelled `A`.

Why other generators do not need this step. The corrections of Section 3.3 are specific to a model that emits a complete Wyckoff string as free text. CrystalFormer predicts only the free coordinates of each site, so the multiplicities and pinned coordinates are correct by construction, and WyFormer predicts no coordinates at all and places atoms with PyXtal before relaxation. Crystal-Text-LLM needs no such parser either, since its string lists every atom explicitly and maps directly to a structure. Among LLM-based Wyckoff generators, PLaID++ resamples strings from which a CIF cannot be parsed, as we do, but reports no coordinate or multiplicity repair.

H.2 PARSER CORRECTIONS IN PRACTICE

Table 10: How often the parser of Section 3.3 intervenes, for 2,500 accepted samples per model (temperature 0.7). "Rejected" counts strings discarded before reaching 2,500 (hard checks or `pymatgen` structure construction). Site-level counts are over all parsed Wyckoff sites; C1 and C2 refer to the multiplicity and fixed-coordinate corrections. The corrections are rarely needed for any SFT variant and never for the RL policies, whose outputs are grammatically consistent.

Model	Rejected	Sites	Invalid letter	C1 triggered	C2 triggered
CrysTune^{gen}	8 (0.3%)	9,229	20 (0.22%)	11 (0.12%)	28 (0.30%)
CrysTune$^{gen+lookup}$	10 (0.4%)	9,419	9 (0.10%)	23 (0.24%)	38 (0.40%)
CrysTune^{SFT}	27 (1.1%)	9,245	8 (0.09%)	16 (0.17%)	45 (0.49%)
CrysTune$^{RL\text{-}naive}$	0	8,622	0	0	0
CrysTune^{RL}	22 (0.9%)	8,464	3 (0.04%)	18 (0.21%)	28 (0.33%)

For every SFT variant, fewer than 0.5% of sites need either correction and about 1% of strings are rejected, so the parser's contribution to the validity reported in Table 2 is small. The auxiliary tasks reduce invalid Wyckoff letters (20 to 8) but not multiplicity mismatches, which are already rare after generation-only training. The preliminary RL policy produces no violations at all, consistent with its collapse onto a few memorized prototypes.

H.3 AUXILIARY-TASK ABLATION

Table 11: Ablation of the SFT tasks (Qwen-3.5-9B, four epochs, 2,500 samples, same protocol as Table 2). *gen*: generation only; *gen+lookup*: generation and Wyckoff lookup; CrysTune^{SFT}: all three tasks. Best value per row in **bold**.

Metric	CrysTune^{gen}	CrysTune$^{gen+lookup}$	CrysTune^{SFT}
Validity (%) ↑	**86.44**	84.12	84.08
Uniqueness (%) ↑	94.77	95.44	**95.72**
Novelty (%) ↑	30.08	33.19	**36.58**
Stability (%) ↑	**1.99**	1.90	1.62
SUN (%) / count ↑	0.37 / **8**	0.29 / 6	**0.38 / 8**
MetaSUN (%) ↑	**3.61**	2.76	2.81
Element diversity ↑	**0.737**	0.720	0.720
Space-group diversity ↑	**0.330**	0.313	0.309
Site-number diversity ↑	14.84	14.48	**15.20**
Physical-size diversity ↑	0.260	**0.262**	**0.262**

Generation-only training already benefits from the Wyckoff representation and reaches the same SUN count as the full curriculum with higher validity and stability. Adding the lookup and canonicalization tasks increases novelty (30.1 to 36.6%) and site-number diversity and reduces stability and MetaSUN slightly. Under these metrics the auxiliary tasks are therefore neutral for generation; their measurable effect is on Wyckoff-table recall (Section G). We keep them in CrysTune^{SFT} because a model that knows the tables is less dependent on the parser's corrections and because the canonicalization task defines the convention the parser enforces, but we do not claim a generation benefit.

H.4 RL VARIANTS

Table 12: GRPO variants evaluated under the protocol of Table 2 (2,500 samples, temperature 0.7). "Legacy" rewards use MACE-MPA, the Materials Project hull, and the MP-20 novelty reference; "bounded" adds the $d/(2+d)$ creativity term, KL 10^{-2}, and entropy 10^{-3}; "buffer" adds the 1,000-structure replay buffer; "OMAT/Alex" switches to MACE-OMAT with the LeMat-Bulk hull and the Alexandria+MP reference. LR is the actor learning rate; FT is full or LoRA fine-tuning. Rows are not single-factor ablations: the buffer runs also lowered the learning rate, and the OMAT/Alex run uses a later checkpoint.

Reward	LR	FT	Step	Valid	Unique	Novel	Stable	SUN	MetaSUN	Elem.	SG
CrysTune$^{\text{SFT}}$ (initialization)				84.08	95.72	36.58	1.62	0.38	2.81	0.720	0.309
legacy (CrysTune$^{\text{RL-naive}}$)	10^{-6}	full	120	99.24	66.51	14.79	0.20	0.04	0.48	0.669	0.017
legacy	10^{-6}	full	180	99.16	56.15	14.60	0.20	0.04	0.52	0.653	0.013
legacy + bounded	10^{-6}	full	120	98.60	84.62	13.63	0.41	0.04	1.78	0.695	0.039
legacy + bounded + buffer	3×10^{-7}	full	120	86.24	95.69	34.51	1.30	0.23	3.25	0.720	0.304
legacy + bounded + buffer	3×10^{-7}	full	180	86.00	95.91	34.70	1.30	0.23	3.02	0.712	0.304
legacy + bounded + buffer	3×10^{-7}	full	240	86.48	95.65	33.95	1.34	0.19	3.42	0.720	0.291
legacy + bounded + buffer	6×10^{-7}	full	240	94.60	94.42	24.82	0.97	0.04	2.71	0.720	0.200
legacy + bounded + buffer	3×10^{-7}	LoRA	120	84.00	95.90	35.43	1.57	0.38	3.38	0.720	0.322
legacy + bounded + buffer	3×10^{-7}	LoRA	240	84.24	96.01	35.47	1.57	0.43	3.09	0.720	0.304
legacy + bounded + buffer	6×10^{-7}	LoRA	180	86.36	95.88	35.39	1.34	0.23	3.33	0.720	0.304
legacy + bounded + buffer	6×10^{-7}	LoRA	240	86.56	95.43	34.57	0.92	0.14	3.14	0.712	0.300
OMAT/Alex + bounded + buffer (CrysTune$^{\text{RL}}$)	6×10^{-7}	full	300	90.04	96.05	34.61	1.69	0.36	3.55	0.729	0.257

Mode collapse under the preliminary recipe. CrysTune$^{\text{RL-naive}}$ illustrates the failure mode that motivated the changes in Section 3.4. After 120 steps of the preliminary recipe, validity rises from 84% to 99% and the mean energy above hull drops from 0.49 to 0.14 eV/atom, which would look like a clear success if these were the only metrics reported. At the same time, uniqueness falls from 96% to 67%, novelty from 37% to 15%, space-group coverage from 0.31 to 0.017 (the policy emits almost only space group 225), and the SUN count from 8 to 1. The collapse was invisible to the reward: shuffling the 2,500 step-180 samples into batches of 320 gives 86% within-batch uniqueness although only 56% of the samples are unique overall, because the policy cycles through a small set of known low-energy prototypes that rarely repeat within a batch but repeat across batches. Continuing to step 180 made uniqueness worse (56%).

Three observations follow from Table 12. First, at a matched step (120), bounding the creativity term and strengthening KL and entropy regularization delays but does not prevent collapse: uniqueness recovers to 85% but novelty and space-group coverage remain at the collapsed level. Second, the replay buffer is the decisive change: with it, uniqueness, novelty, and space-group coverage return to the SFT level for both full and LoRA fine-tuning and for both learning rates at 3×10^{-7}, at the cost of the validity gain. Third, at the higher learning rate the full-parameter policy drifts faster (validity 94.6%, novelty 24.8% at step 240), whereas the LoRA policy stays close to SFT; the OMAT/Alex run at the same learning rate (evaluated at step 300, so not a matched-step comparison) shows that a better-matched energy model and larger novelty reference recover the validity gain while keeping novelty at 34.6%. Checkpoint choice should be driven by SUN and MetaSUN counts together with coverage, not by training reward, which increases monotonically in every run.

H.5 DECODING SENSITIVITY

Table 13: Effect of decoding on the step-240 buffer variants (2,500 samples each). Each cell reports temperature 0.7 / top-p 1.0 (default), temperature 1.0 / top-p 0.7, and temperature 1.0 / top-p 1.0.

Model (step 240)	Valid	Novel	Unique	Stable	SUN (count)	MetaSUN
buffer, full, 3×10^{-7}	86.5 / 91.3 / 78.1	34.0 / 23.5 / 48.9	95.7 / 88.0 / 98.8	1.34 / 2.41 / 1.02	0.19 (4) / 0.22 (5) / 0.15 (3)	3.42 / 2.89 / 2.36
buffer, LoRA, 3×10^{-7}	84.2 / 90.0 / 75.0	35.5 / 23.9 / 51.6	96.0 / 88.2 / 99.0	1.57 / 2.53 / 0.85	0.43 (9) / 0.13 (3) / 0.16 (3)	3.09 / 2.58 / 2.67
buffer, full, 6×10^{-7}	94.6 / 96.8 / 88.9	24.8 / 20.0 / 36.7	94.4 / 85.7 / 98.7	0.97 / 1.20 / 1.03	0.04 (1) / 0.08 (2) / 0.14 (3)	2.71 / 2.27 / 2.20
buffer, LoRA, 6×10^{-7}	86.6 / 91.1 / 77.4	34.6 / 24.3 / 50.0	95.4 / 88.1 / 99.1	0.92 / 2.11 / 0.88	0.14 (3) / 0.22 (5) / 0.10 (2)	3.14 / 2.59 / 2.69

Decoding moves validity, novelty, and uniqueness by 10–15 percentage points, which is as large as the differences between RL variants, while SUN counts change by a few structures in either direction. Nucleus sampling at top-$p = 0.7$ raises validity and stability and halves novelty; unrestricted sampling at temperature 1 does the opposite. All comparisons in the main text use the same decoding for every LLM model, and we recommend that decoding be reported alongside any LLM crystal-generation result. A finer sweep over temperature and top-p on 500 samples of CrysTune$^{\text{SFT}}$ showed the same monotone trade-off.

H.6 Novelty against the training set

Table 14: The same 2,500 structures per model evaluated with the MP-20 training set (27,136 structures) instead of LeMat-Bulk (5.3M structures) as the novelty and SUN reference. Validity, uniqueness, and stability are unchanged from Table 2.

Model	Novelty (MP-20)	Novelty (LeMat-Bulk)	SUN (MP-20)	SUN (LeMat-Bulk)
Chemeleon2	98.16	76.54	0.00	0.00
WyFormer	94.65	79.46	0.00	0.00
DiffCSP	89.95	68.96	0.42	0.21
MatterGen	91.84	73.44	0.25	0.12
CrystalFormer	73.98	43.05	0.29	0.06
CrysTune$^{\text{SFT}}$	74.02	36.58	0.81	0.38
CrysTune$^{\text{RL}}$	77.08	34.61	0.89	0.36
CrysTune$^{\text{RL-naive}}$	89.68	14.79	0.16	0.04

Against the training set alone, every model looks far more novel, and the collapsed CrysTune$^{\text{RL-naive}}$ policy appears more novel than CrysTune$^{\text{SFT}}$, because the prototypes it repeats are present in LeMat-Bulk but not in MP-20. Under the MP-20 reference, CrysTune$^{\text{RL}}$ has 20 SUN structures (0.89%) against 17 for CrysTune$^{\text{SFT}}$ (0.81%), whereas under LeMat-Bulk both have 8; the RL policy's additional stable structures are novel with respect to its training data but known to the larger database. On the same 320 structures, an MP-20 reference labelled 299 as novel and LeMat-Bulk labelled 52. This is why we moved the RL novelty reference to Alexandria+MP and why we report LeMat-Bulk novelty in the main text.

H.7 SPACE-GROUP COVERAGE

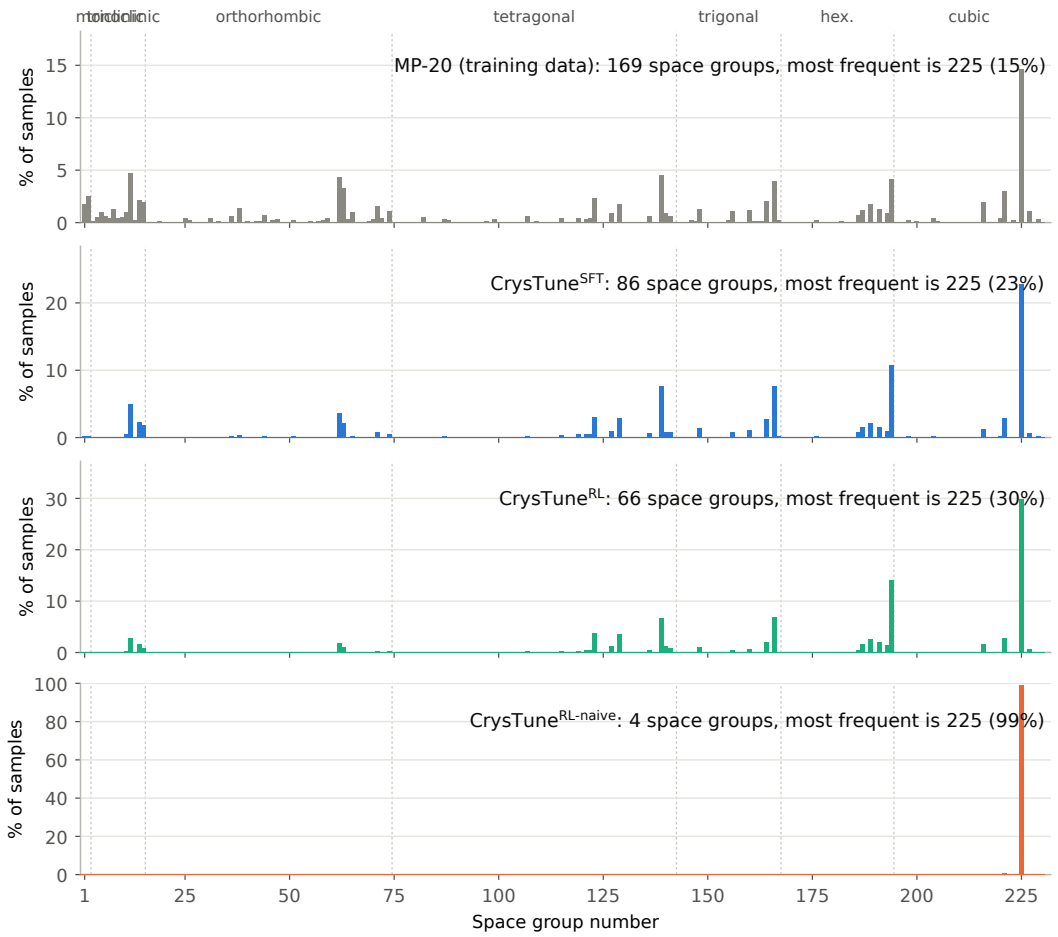


Figure 4: Space-group distribution of 2,500 samples from each model, with the MP-20 training data for reference; dashed lines separate the seven crystal systems. CrysTune$^{\text{SFT}}$ covers 86 of the 230 space groups and CrysTune$^{\text{RL}}$ 66, both inheriting the training data's preference for the cubic space group 225 and sharpening it (15% of MP-20 crystals, 23% of CrysTune$^{\text{SFT}}$ samples, 30% of CrysTune$^{\text{RL}}$ samples). The policy trained without the mitigation strategies of Section 3.4 emits only four space groups, with 99% of its samples in space group 225.

H.8 QUALITATIVE ANALYSIS OF STABLE AND METASTABLE STRUCTURES

Figures 5 and 6 show metastable structures chosen for chemical and symmetry variety. The structures that ORB-v3 places on or below the hull are far less varied, and examining them exposes a limitation of strict-threshold stability metrics computed with a single potential.

Table 15: Chemistry of the samples that ORB-v3 places on or below the hull (stable, $\Delta E_{\text{hull}} \leq 0$) and between the hull and 0.1 eV/atom (metastable), out of 2,500 generated structures per model.

Model	Subset	Count	Contains Yb	Contains a lanthanide
CrysTune$^{\text{SFT}}$	All samples	2,500	93 (3.7%)	1,019 (41%)
	Metastable	1,027	51 (5.0%)	612 (60%)
	Stable	34	17 (50%)	25 (74%)
	SUN	8	7 (88%)	8 (100%)
CrysTune$^{\text{RL}}$	All samples	2,500	114 (4.6%)	944 (38%)
	Metastable	1,093	57 (5.2%)	593 (54%)
	Stable	38	22 (58%)	32 (84%)
	SUN	8	8 (100%)	8 (100%)

Ytterbium appears in under 5% of the generated samples of either model but in half or more of the structures called stable, and in seven of the eight SUN structures of CrysTune$^{\text{SFT}}$ and all eight of CrysTune$^{\text{RL}}$; lanthanides in general are enriched from about 40% of samples to 74–84% of the stable subsets. The metastable subsets show no comparable enrichment. That both the supervised and the RL policy show the same pattern, while their chemistry otherwise differs, indicates a property of the evaluation rather than of a particular model. Two causes are likely. First, machine-learned interatomic potentials are least accurate for lanthanides, whose f electrons are poorly described by the density-functional data they are trained on, and Yb is a well-known difficult case. Second, the convex hull is sparse for rare-earth intermetallics, so a moderate error in the predicted energy is enough to place a structure below a hull defined by few competing phases. We therefore do not interpret these structures as discoveries. More generally, a stable count of a few dozen out of 2,500, concentrated in one chemical family, is a fragile basis for ranking models, which is consistent with the protocol dependence documented in Section H.9. The metastable subsets, two orders of magnitude larger and chemically representative of the generated distribution, are the more reliable signal, and this is why we report metastable-SUN alongside SUN throughout.

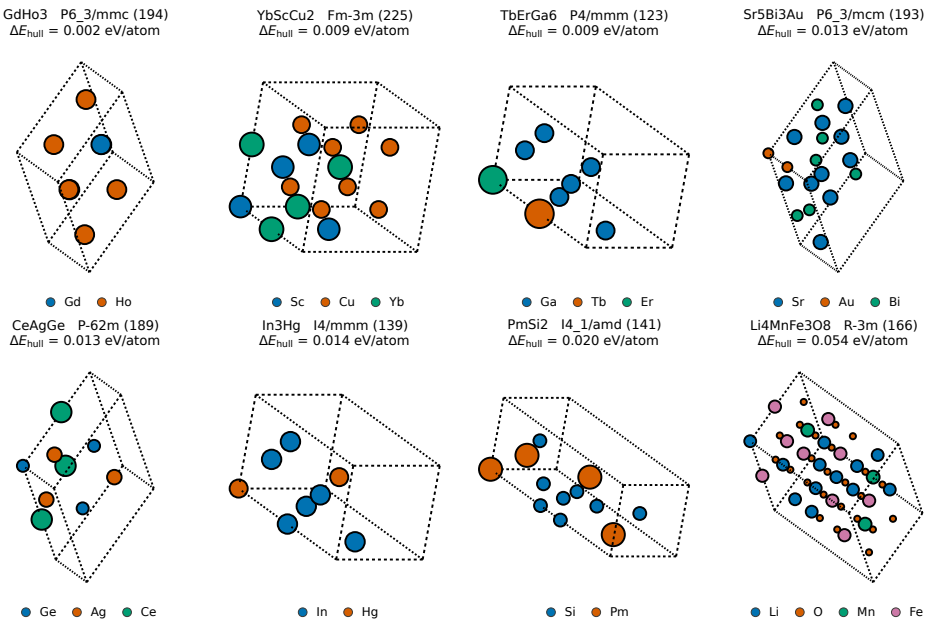


Figure 5: Eight metastable, unique, and novel structures generated by CrysTune$^{\text{RL}}$, selected from the 80 such structures for chemical and symmetry variety. Atom colours are assigned per panel and given in the legend below it.

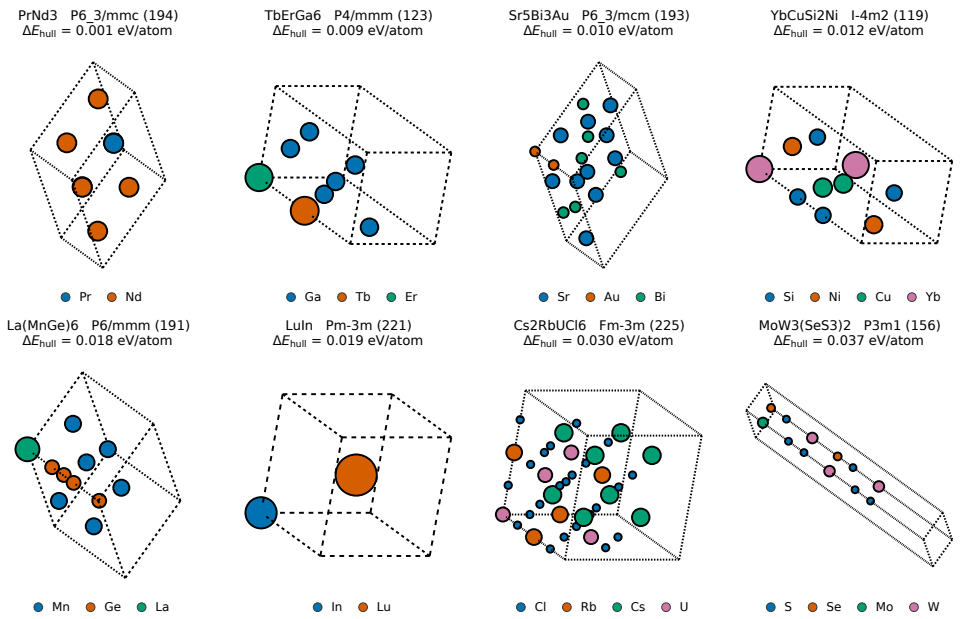


Figure 6: Eight metastable, unique, and novel structures generated by CrysTune$^{\text{SFT}}$, selected from the 59 such structures for chemical and symmetry variety; the counterpart of Figure 5 for the supervised model. Atom colours are assigned per panel and given in the legend below it.

H.9 Evaluation under alternative protocols

We re-evaluated the same 2,500 structures per model with two further pipelines. **MatterGen** (Zeni et al., 2023): relaxation with MatterSim, energies with MP2020 corrections against the MatterGen reference hull, novelty and uniqueness against the Alexandria+MP reference, stable meaning $\Delta E_{\text{hull}} \leq 0.1$ eV/atom. **CHGNet**: the relaxation-and-metrics code released with FlowMM (Miller et al., 2024), with the default pretrained CHGNet checkpoint, relaxation of at most 500 steps, a Materials Project phase diagram from February 2023, novelty against the MP-20 training and validation splits with default `StructureMatcher` settings, and single-element structures excluded; metastable and stable use thresholds of 0.1 and 0 eV/atom. In both pipelines rates are over all 2,500 generated structures, not over valid ones.

Table 16: Stability and SUN (%) of the same 2,500 structures per model under three evaluation protocols. LeMat-GenBench rates are over valid structures (Table 2); MatterGen and CHGNet rates are over all 2,500. – : not yet evaluated.

	LeMat-GenBench (ORB-v3)		*MatterGen (MatterSim)*		*CHGNet*			
Model	Stable	SUN	$\text{Stable}_{0.1}$	$\text{SUN}_{0.1}$	Metast.	MetaSUN	Stable	SUN
Chemeleon2	0.00	0.00	67.40	30.88	80.16	70.72	21.04	17.84
DiffCSP	0.59	0.21	37.56	14.32	36.04	31.36	10.76	8.48
MatterGen	0.42	0.12	43.64	22.40	62.40	54.80	9.52	8.28
CrystalFormer	0.58	0.06	40.80	7.52	52.96	33.68	12.52	8.40
WyFormer	0.05	0.00	–	–	32.32	26.40	6.84	5.28
CrysTxtLLM	0.46	0.10	37.04	5.20	–	–	–	–
CrysTune$^{\text{SFT}}$	1.62	0.38	52.56	7.20	64.84	40.96	18.96	11.24
CrysTune$^{\text{RL-naive}}$	0.20	0.04	61.12	4.44	–	–	–	–
CrysTune$^{\text{RL}}$	1.69	0.36	49.56	7.08	–	–	–	–

The absolute numbers and the ranking of the strongest models depend strongly on the protocol. Chemeleon2 has no stable structure under the LeMat-Bulk hull but the highest SUN rate under both the 2023 Materials Project hull with CHGNet (17.8%) and the MatterGen protocol with MatterSim (30.9%), and MatterGen, which is fourth under LeMat-GenBench, is second under its own protocol

(22.4%). Under the MatterGen protocol CrysTune ranks below Chemeleon2, MatterGen, DiffCSP, and CrystalFormer on SUN (7.1–7.2% against 7.5–30.9%), whereas under LeMat-GenBench it ranks first and under CHGNet second. The LeMat-Bulk hull, which includes Alexandria and OQMD entries, lies below the Materials Project hull for many chemistries, so structures close to known prototypes that are stable with respect to Materials Project are often above it; conversely, the lenient 0.1 eV/atom threshold of the MatterGen protocol rewards models that generate many near-hull structures, which favours the diffusion models. In contrast, the comparisons within this paper are consistent across protocols: CrysTune$^{\text{SFT}}$ is more stable and has more SUN structures than CrysTxtLLM under both protocols where both were evaluated (7.20% against 5.20% SUN under MatterGen), the collapsed CrysTune$^{\text{RL-naive}}$ policy has a lower SUN rate than CrysTune$^{\text{SFT}}$ despite a higher stable rate, and CrysTune$^{\text{RL}}$ is close to CrysTune$^{\text{SFT}}$ under MatterGen as well (7.08% against 7.20% SUN, with a lower mean energy above hull of 0.205 against 0.246 eV/atom and a higher novelty of 50.4% against 45.5% against the Alexandria+MP reference). Under the lenient CHGNet and MatterGen thresholds, CrysTune is below the best diffusion models.

H.10 Crystal-system-conditioned generation

Table 17: Number (and fraction) of the 2,500 generations per requested crystal system that are valid and match the request, for the three Qwen-3.5-9B models of Figure 2b. The conditional policy is evaluated at 200 GRPO steps.

Requested system	CrysTune$^{\text{SFT}}$	CrysTune$^{\text{RL}}$	CrysTune$^{\text{RL}}{}_{\text{CS}}$
Cubic	2,404 (96.2%)	2,402 (96.1%)	2,413 (96.5%)
Hexagonal	2,243 (89.7%)	2,293 (91.7%)	2,367 (94.7%)
Tetragonal	2,129 (85.2%)	2,170 (86.8%)	2,260 (90.4%)
Trigonal	1,923 (76.9%)	2,017 (80.7%)	2,068 (82.7%)
Orthorhombic	1,804 (72.2%)	1,872 (74.9%)	2,004 (80.2%)
Monoclinic	976 (39.0%)	1,120 (44.8%)	1,122 (44.9%)
Triclinic	317 (12.7%)	429 (17.2%)	416 (16.6%)
All seven (17,500)	11,796 (67.4%)	12,303 (70.3%)	12,650 (72.3%)

I Wyckoff Tables: 174 and 167

Table 18: Wyckoff positions for space group 174 ($P\bar{6}$).

Mult.	Letter	Site sym.	Coordinates
1	a	$\bar{6}$ ..	$(0,\ 0,\ 0)$
1	b	$\bar{6}$ ..	$(0,\ 0,\ 1/2)$
1	c	$\bar{6}$ ..	$(1/3,\ 2/3,\ 0)$
1	d	$\bar{6}$ ..	$(1/3,\ 2/3,\ 1/2)$
1	e	$\bar{6}$ ..	$(2/3,\ 1/3,\ 0)$
1	f	$\bar{6}$ ..	$(2/3,\ 1/3,\ 1/2)$
2	g	3 ..	$(0,\ 0,\ z),\ (0,\ 0,\ -z)$
2	h	3 ..	$(1/3,\ 2/3,\ z),\ (1/3,\ 2/3,\ -z)$
2	i	3 ..	$(2/3,\ 1/3,\ z),\ (2/3,\ 1/3,\ -z)$
3	j	m ..	$(x,\ y,\ 0),\ (-y,\ x-y,\ 0),$ $(-x+y,\ -x,\ 0)$
3	k	m ..	$(x,\ y,\ 1/2),\ (-y,\ x-y,\ 1/2),$ $(-x+y,\ -x,\ 1/2)$
6	l	1	$(x,\ y,\ z),\ (-y,\ x-y,\ z),$ $(-x+y,\ -x,\ z),\ (x,\ y,\ -z),$ $(-y,\ x-y,\ -z),\ (-x+y,\ -x,\ -z)$

Table 19: Wyckoff positions for space group 167 ($R\bar{3}c$, hexagonal axes). Unlike the primitive group in Table 18, the rhombohedral lattice carries three centering operations (top row): each listed coordinate is replicated by adding every centering vector, tripling the row count to give the full multiplicity. The $12c$ row therefore expands to $4 \times 3 = 12$ atoms in the unit cell.

Mult.	Letter	Site sym.	Coordinates (per centering block)
		Centering vectors:	$(0,\ 0,\ 0)+\quad (\frac{2}{3},\ \frac{1}{3},\ \frac{1}{3})+\quad (\frac{1}{3},\ \frac{2}{3},\ \frac{2}{3})+$
6	a	32	$(0,\ 0,\ 1/4),\ (0,\ 0,\ 3/4)$
6	b	$\bar{3}$	$(0,\ 0,\ 0),\ (0,\ 0,\ 1/2)$
12	c	3 ..	$(0,\ 0,\ z),\ (0,\ 0,\ -z+1/2),$ $(0,\ 0,\ -z),\ (0,\ 0,\ z+1/2)$
18	d	$\bar{1}$	$(1/2,\ 0,\ 0),\ (0,\ 1/2,\ 0),$ $(1/2,\ 1/2,\ 0),\ (0,\ 1/2,\ 1/2),$ $(1/2,\ 0,\ 1/2),\ (1/2,\ 1/2,\ 1/2)$
18	e	.2	$(x,\ 0,\ 1/4),\ (0,\ x,\ 1/4),$ $(-x,\ -x,\ 1/4),\ (-x,\ 0,\ 3/4),$ $(0,\ -x,\ 3/4),\ (x,\ x,\ 3/4)$
36	f	1	$(x,\ y,\ z),\ (-y,\ x-y,\ z),$ $(-x+y,\ -x,\ z),\ (y,\ x,\ -z+1/2),$ $(x-y,\ -y,\ -z+1/2),$ $(-x,\ -x+y,\ -z+1/2),$ $(-x,\ -y,\ -z),\ (y,\ -x+y,\ -z),$ $(x-y,\ x,\ -z),\ (-y,\ -x,\ z+1/2),$ $(-x+y,\ y,\ z+1/2),$ $(x,\ x-y,\ z+1/2)$

J Inference Projection

Tag	Failure mode
`parsing_error`	Syntax, space group, Wyckoff symbol, or numeric field (Section H.1)
`comp_invalid`	Element symbol $\hat{e}_i$ unrecognized by pymatgen
`max_atoms`	Any $Z(\hat{e}_i) > 94$
`valid`	Parser returns $\mathcal{P}(\hat{s}) \in \mathcal{S}$

For ease of reimplementation, we restate the C2 and C3 operations of Section 3.3 in compact symbolic form.

Free-axis set. For each (n_G, ℓ), let $\mathcal{F}(n_G, \ell) \subseteq \{1, 2, 3\}$ be the indices of axes that appear as free parameters in $\tau(n_G, \ell)$; the complement $\{1, 2, 3\} \setminus \mathcal{F}(n_G, \ell)$ is the set of pinned axes. For example, $\mathcal{F}(174, 3j) = \{1, 2\}$ (axes x and y free) and $\mathcal{F}(167, 6a) = \varnothing$ (all axes pinned).

C2 (fixed-coordinate snap). For each site i,

$$\hat{r}_{i,a} \ \leftarrow \ \tau(\hat{n}_G, \hat{\ell}_i)_a \quad \text{for every } a \notin \mathcal{F}(\hat{n}_G, \hat{\ell}_i), \tag{8}$$

leaving free axes unchanged. After this update, the pinned components of $\hat{\mathbf{r}}_i$ exactly match the site template.

Site-specific Wyckoff operations. For Wyckoff letter ℓ in space group n_G, the International Tables list $m = \mathrm{Mult}(n_G, \ell)$ affine maps $W_1^{(\ell)}, \ldots, W_m^{(\ell)}$, each of the form $W_j\mathbf{r} \ = \ \mathbf{R}_j\mathbf{r} + \mathbf{t}_j \ (\mathrm{mod}\ \mathbf{1})$, following the same convention as the elements of G in Section 2; for centered lattices, the centering vectors are folded into the $\mathbf{t}_j$. The rows of Tables 18 and 19 display these maps as coordinate expressions for two example space groups.

C3 (orbit expansion). The m_i atomic positions of site i are

$$\mathcal{O}_i \ = \ \left\{ W_j^{(\hat{\ell}_i)}\, \hat{\mathbf{r}}_i \ : \ j = 1, \ldots, m_i \right\}. \tag{9}$$

The CIF is built from $\bigcup_i \mathcal{O}_i$, paired with the corresponding species $\hat{e}_i$.

K LeMat-GenBench Metrics

We evaluate each generated crystal after parsing it as a `pymatgen` structure. Ratio-based quantities are reported as percentages, while energy and displacement quantities are reported in their native units. Although the benchmark configuration supports multiple machine-learned interatomic potentials (MLIPs), all stability-related values reported here use the ORB-v3 model only, specifically the ORB-specific convex hull results, rather than an ensemble average.

Sampling and parser rejections. Structures rejected by the parser (Section H.1) are resampled until 2,500 are obtained; the rejection rate is a property of the model and was 1.1% for CrysTune$^{\text{SFT}}$ and 0.9% for CrysTune$^{\text{RL}}$. The novelty reference used here is two orders of magnitude larger than MP-20 and six times larger than the Alexandria+MP reference used in the reward, so novelty under this protocol is a stricter criterion than novelty during training.

Validity. `overall_validity` is the percentage of generated structures that pass all validity checks. A structure is counted as valid only if it is charge neutral within the configured tolerance, has no unrealistically short interatomic distances, and passes basic physical plausibility checks on density, lattice parameters, CIF serialization, and space-group assignment.

Uniqueness and novelty. `overall_uniqueness` is the percentage of valid structures that are not duplicates of another generated structure, and `overall_novelty` is the percentage of valid structures not matched to the LeMat-Bulk `compatible_pbe` training split (5,335,299 structures). Both use `StructureMatcher` after a composition pre-filter. We note that the benchmark also offers a BAWL hash fingerprint; on the same 2,500 CrysTune$^{\text{SFT}}$ structures it reports 90.1% novelty instead of 36.6%, so the fingerprint must be stated with every number.

Stability and MetaSUN. For each structure, ORB-v3 is used to compute the formation energy and the energy above the ORB convex hull after relaxation (at most 50 steps, $f_{\max} = 0.02$ eV/Å). `stability_score` is the percentage of valid structures with energy above hull at most 0 eV/atom, and `metastability_score` the percentage with at most 0.1 eV/atom. The mean ORB energy above hull is reported as `e_above_hull` in eV/atom.

SUN and MetaSUN. `sun_score` is the percentage of valid structures that are simultaneously stable, unique, and novel: the stable subset is deduplicated, and the remaining structures are tested for novelty. `msun_score` applies the same criterion to the metastable subset (above 0 and at most 0.1 eV/atom). Because the denominator is the number of valid structures, two models with the same SUN count can have different SUN rates; we therefore also report counts.

Relaxation and diversity. `relaxation_rmse` is the mean root-mean-square Cartesian displacement between the input structure and the ORB-relaxed structure, in Å. Element diversity measures coverage of chemical elements, space-group diversity measures coverage of space groups, site-number diversity is the Vendi score of the number of sites per cell, and physical-size diversity summarizes variation in density, lattice parameters, volume, and packing factor.

L Compute

All experiments used nodes with four NVIDIA H100 80GB GPUs. SFT of Qwen-3.5-9B took about 5.5 hours for four epochs. Each GRPO step takes 150–170 s after a 20–50 minute cold start, so 320 steps take about 15 hours; the energy evaluation, replay-buffer matching, and rollout generation dominate, and LoRA does not materially reduce wall time. Generating 2,500 parseable structures takes about 30 minutes, and one LeMat-GenBench evaluation of 2,500 structures takes 30–70 minutes on four L40S GPUs.