FAMILYSPACE: FROM EMERGENT CHEMICAL SEMANTICS TO CONTROLLABLE CRYSTAL GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Latent crystal generative models typically combine a variational autoencoder (VAE) for learning atom-wise representations with a flow model that transports a simple source distribution to the learned latent distribution. However, previous studies have mainly focused on the quality of generated crystals, leaving the structure and generative utility of the atom-wise latent space largely unexplored. In this work, we show that even a vanilla crystal VAE implicitly organizes atom-wise latent representations according to elemental families, with different families occupying largely separated regions of the latent space. We term this phenomenon *emergent chemical semantics*. This structure suggests that sampling from a family-specific latent region can bias the decoder toward atoms from the corresponding elemental family. Building on this insight, we propose FamilySpace, a framework for controlling the elemental-family composition of generated crystals. FamilySpace assigns each elemental family a distinct source component, forming a family-indexed multi-source distribution, and trains a flow model using source-target pairs matched by family. In inference, a user-specified family template selects the source component for each atom-wise representation, enabling direct control over the family composition of generated crystals. Experiments on MP-20 and MPTS-52 show that FamilySpace achieves competitive crystal-generation performance while matching prescribed elemental-family compositions. Moreover, selected family templates increase the fraction of generated crystals predicted to be stable or metastable, demonstrating that latent-space composition control can support targeted exploration of promising chemical regions without additional optimization techniques such as reinforcement learning. Code is available at `https://anonymous.4open.science/r/FamilySpace-3E7E/`

1 INTRODUCTION

Crystalline materials discovery underpins advances in energy, electronics, catalysis, and many other technologies (Butler et al., 2018; Wang et al., 2026; Cheng et al., 2026). Discovering such materials requires navigating an enormous chemical design space spanning diverse compositions, atomic arrangements, and lattice geometries. Conventional discovery workflows rely heavily on iterative experiments and computationally intensive first-principles calculations, limiting the scale and speed of materials exploration (Potyrailo et al., 2011; Maier, 2019; Jain et al., 2011). To accelerate this process, deep generative models have emerged as a promising approach for learning distributions over crystal structures and directly proposing previously unseen candidates (Xie et al., 2022; Luo et al., 2023; Zhao et al., 2021). In particular, diffusion- and flow-based models have further advanced crystal generation by jointly modeling atomic species, fractional coordinates, and lattice parameters (Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024; Zeni et al., 2025).

Building on recent advances in latent generative modeling for molecules (Xu et al., 2023; Wang et al., 2024; Bian et al., 2024; Ding & Hofmann, 2025; Zhang et al., 2025; Slautin et al., 2026; Luo et al., 2025b), recent work has explored latent generative models for improving crystal generation performance (Khastagir et al., 2025a; Joshi et al., 2025). In a standard flow-based framework (Joshi et al., 2025), a variational autoencoder (VAE) first learns continuous representations of crystals through encoding and reconstruction. A separate flow model then maps an initial source distribution to the distribution of VAE-encoded representations. During inference, the VAE decoder converts representations transported by the flow into crystal structures. Despite the central role of these rep-

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.

resentations in generation, existing work primarily focuses on the quality of generated crystals, with limited attention to the organization of the latent space itself. Understanding how these representations are organized could inform the design of latent generative models and sampling strategies.

We therefore ask: *What organization emerges in the latent space of a crystal VAE, and how can it guide generation?* Recent work has shown that atomic representations learned by machine-learning interatomic potentials can reflect consistent geometric organization of atomic environments. In particular, representations of atoms from the same periodic-table group tend to cluster together (Li & Walsh, 2026). This motivates us to examine whether similar geometric organization can emerge when learning latent representations of crystals. To this end, we analyze a vanilla atom-wise crystal VAE trained only with reconstruction loss and Kullback–Leibler (KL) regularization.

We find that its atomic latent representations naturally organize by elemental family, despite the absence of family labels during VAE training. Here, a family refers to a periodic-table category such as halogens or transition metals, following the mapping in Appendix A.1. As shown in Figure 1, representations from the same family form coherent neighborhoods, while representations from different families occupy largely non-overlapping regions. We refer to this geometric organization as *emergent chemical semantics*. This finding suggests that latent representations sampled from a family-specific region may preferentially decode into atoms from the corresponding family, providing a potential basis for controlling elemental family composition. However, the standard latent flow models (Joshi et al., 2025) do not make full use of such emergent chemical semantics. The challenge is therefore to determine how to incorporate it into latent flow modeling to guide generation.

To bridge this gap, we introduce **FamilySpace**, which transforms the observed semantics into a family-indexed multi-source distribution for latent flow matching. For each elemental family, FamilySpace constructs a local source component from the corresponding atomic representations in the pretrained VAE posterior. A single shared flow is trained by pairing samples from each source component with target atomic latents from the same family, using the standard flow-matching objective. The VAE remains frozen, and family labels are used for source construction and source–target pairing. Figure 2 illustrates the overall framework. At inference, a family template specifies the desired elemental families and their proportions, determining the source component for each atomic representation. Source selection therefore controls family composition while leaving exact element identities and crystal geometry to be generated by the model.

Experiments on MP-20 and MPTS-52 demonstrate competitive generation performance, while control experiments show that generated crystals accurately follow the requested family compositions. Moreover, selected chemically motivated templates increase the fraction of generated crystals predicted to be stable or metastable, without additional model optimization such as reinforcement learning. These results show how emergent chemical semantics can support both controllable generation and targeted materials exploration.

Our main contributions are as follows:

- We identify *emergent chemical semantics* in a vanilla atom-wise crystal VAE. Although trained only with reconstruction and KL regularization, the VAE learns latent representations that naturally organize by elemental family.
- We introduce FamilySpace, which converts this emergent chemical semantics into a family-indexed multi-source distribution for latent flow matching. Pairing source and target latents from the same family enables family-composition control through source selection.
- We demonstrate FamilySpace's precise family-composition control and competitive crystal generation performance. The controllable generation mechanism of FamilySpace enables targeted exploration of chemically meaningful regions of materials space, without additional optimization techniques such as reinforcement learning.

2 Related Work

Generative models for crystalline materials. Recent crystal generative models predominantly model crystal variables directly, jointly generating atomic species, fractional coordinates, and lattice parameters with diffusion or flow-based objectives (Xie et al., 2022; Jiao et al., 2023; Miller et al., 2024; Zeni et al., 2025). Subsequent work has improved scalability, stability, conditioning,

and exploration through stronger priors, symmetry-aware representations, property guidance, and multimodal inputs (Sriram et al., 2024; Cornet et al., 2025; Kazeev et al., 2025; Kelvinius et al., 2025; Höllmer et al., 2025; Das et al., 2025; Levy et al., 2025; Chang et al., 2025; Khastagir et al., 2025b; Seong et al., 2026; Höllmer & Martiniani, 2026). These works have substantially advanced direct generative modeling of explicit crystal variables.

Latent generative modeling for crystalline materials. Latent crystal generative models separate representation learning from generative distribution modeling. CrysLDM performs diffusion over crystal representations learned by a VAE (Khastagir et al., 2025a), while ADiT combines a shared all-atom VAE with latent flow matching to jointly model periodic crystals and non-periodic molecules (Joshi et al., 2025). In these models, the VAE learns to encode and reconstruct crystal structures, whereas the diffusion or flow model learns the distribution of the resulting latent representations. In a related direction, Chemeleon learns aligned text and crystal representations for text-guided crystal generation (Park et al., 2025). Chemeleon2 further introduces a continuous VAE latent space, performing denoising diffusion in this space and incorporating reinforcement learning for materials exploration (Park & Walsh, 2026).

3 PRELIMINARIES

3.1 CRYSTAL REPRESENTATION AND ATOM-WISE LATENT SPACE

A periodic crystal containing N atoms is represented as $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$, where $\mathbf{A} = (a_1, \ldots, a_N)$ denotes the atomic species, $\mathbf{F} \in [0,1)^{N\times 3}$ contains the fractional coordinates, and $\mathbf{L}$ specifies the periodic unit cell. We parameterize $\mathbf{L}$ by its three cell lengths $\mathbf{l} = (a, b, c)$ and three cell angles $\boldsymbol{\alpha} = (\alpha, \beta, \gamma)$. Following the latent representation proposed by Joshi et al. (2025), we encode a crystal into an atom-wise latent sequence $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_N) \in \mathbb{R}^{N\times d}$, where each latent representation $\mathbf{z}_i$ is associated with one atom.

3.2 ATOM-WISE VAE

Following ADiT (Joshi et al., 2025), the first stage of latent generative modeling maps the crystal variables $\mathcal{M}$ into a continuous atom-wise latent space using a VAE. The encoder produces a Gaussian posterior $q_\phi(\mathbf{z}_i \mid \mathcal{M}) = \mathcal{N}\big(\boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}_i^2)\big)$, from which the atom-wise latent sequence $\mathbf{Z}$ is sampled. The decoder maps $\mathbf{Z}$ back to the crystal variables. For periodic crystals, the reconstruction objective includes atomic species, fractional coordinates, cell lengths, and cell angles: $\mathcal{L}_{\mathrm{rec}} = \lambda_A\mathcal{L}_A + \lambda_F\mathcal{L}_F + \lambda_l\mathcal{L}_l + \lambda_\alpha\mathcal{L}_\alpha$, where $\mathcal{L}_A$ is the cross-entropy loss for atomic species, while $\mathcal{L}_F$, $\mathcal{L}_l$, and $\mathcal{L}_\alpha$ are reconstruction losses for fractional coordinates, cell lengths, and cell angles, respectively. The complete VAE objective is

$$\mathcal{L}_{\mathrm{VAE}} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{KL}} D_{\mathrm{KL}}\left(q_\phi(\mathbf{Z} \mid \mathcal{M}) \,\|\, \mathcal{N}(\mathbf{0}, \mathbf{I})\right). \tag{1}$$

The VAE is trained solely to reconstruct crystal structures and regularize the latent distribution.

3.3 LATENT FLOW MATCHING

After training the VAE, the second stage of latent generative modeling learns to transform a simple source distribution into the distribution of encoded crystal latents. Following ADiT (Joshi et al., 2025), let $\mathbf{Z}^{(1)}$ denote a clean latent representation obtained from the frozen VAE encoder and let $\mathbf{Z}^{(0)} \sim p_0 = \mathcal{N}(\mathbf{0}, \mathbf{I})$ denote a sample from the standard Gaussian source. A linear conditional path between the source and target is defined as

$$\mathbf{Z}^{(t)} = (1-t)\mathbf{Z}^{(0)} + t\mathbf{Z}^{(1)}, \qquad t \sim \mathcal{U}(0,1). \tag{2}$$

The corresponding conditional velocity field is

$$\mathbf{u}_t\left(\mathbf{Z}^{(t)} \mid \mathbf{Z}^{(1)}\right) = \frac{\mathbf{Z}^{(1)} - \mathbf{Z}^{(t)}}{1-t}. \tag{3}$$

ADiT parameterizes this vector field by predicting the clean endpoint from the intermediate latent, $\widehat{\mathbf{Z}}^{(1)} = F_\theta\big(\mathbf{Z}^{(t)}, t\big)$, which induces the predicted velocity $\mathbf{v}_\theta(\mathbf{Z}^{(t)}, t) = \frac{\widehat{\mathbf{Z}}^{(1)} - \mathbf{Z}^{(t)}}{1-t}$. The objective

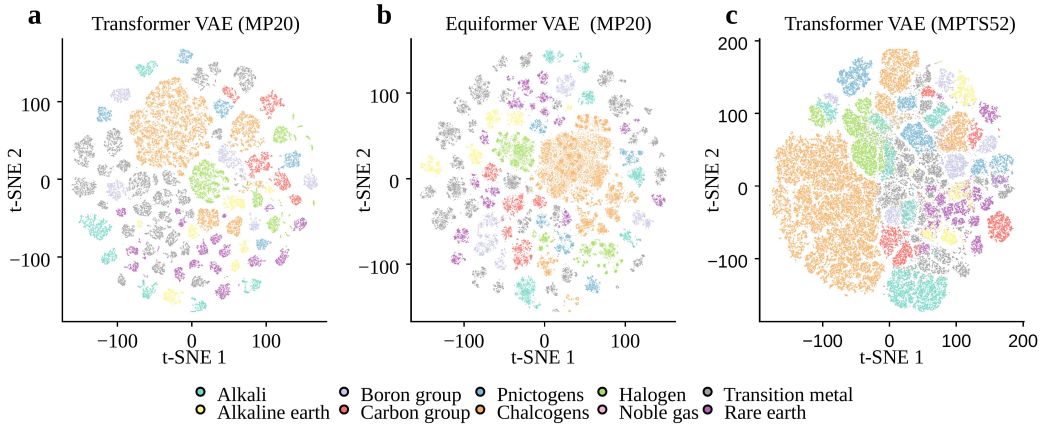


Figure 1: **Emergent chemical semantics in atom-wise latent representations.** We visualize the posterior-mean atom-wise latent representations extracted from the complete test splits using t-SNE. Each point represents an atom and is colored by its periodic-table family only for visualization. (a) Transformer VAE on MP-20. (b) Equiformer VAE on MP-20. (c) Transformer VAE on MPTS-52.

minimizes the discrepancy between the predicted and target vector fields, as follows:

$$\mathcal{L}_{\text{LFM}} = \mathbb{E}_{t,\mathbf{Z}^{(0)},\mathbf{Z}^{(1)}} \left[\frac{1}{(1-t)^2} \left\| \widehat{\mathbf{Z}}^{(1)} - \mathbf{Z}^{(1)} \right\|_F^2 \right]. \tag{4}$$

At inference time, samples are initialized from p_0 and transported from $t = 0$ to $t = 1$ by integrating the learned vector field, after which the frozen VAE decoder reconstructs the terminal latent representation into a crystal.

4 FAMILYSPACE

4.1 EMERGENT CHEMICAL SEMANTICS IN A VANILLA CRYSTAL VAE

We begin with a vanilla atom-wise crystal VAE (Joshi et al., 2025) and examine how its learned atomic representations are organized in latent space. Given a crystal structure, the encoder produces an atom-wise posterior distribution for each atom, from which latent representations are sampled and passed to the decoder for crystal reconstruction. The VAE is trained using only the standard reconstruction loss and KL regularization in Equation 1. We analyze the learned representations using the posterior mean $\boldsymbol{\mu}_i$ of each atom i. As shown in Figure 1, a consistent geometric organization emerges across both Transformer- and Equiformer-based VAEs. Representations of atoms from the same elemental family form coherent neighborhoods, whereas representations from different families occupy largely non-overlapping regions. We refer to this spontaneously learned family-level organization as *emergent chemical semantics*. Further visualizations and analysis are provided in Appendix A.3. These findings suggest that reconstruction-based crystal representation learning naturally gives rise to family-level chemical organization in the latent space. In the following subsection, we investigate how this organization can be exploited to guide generation.

4.2 FAMILY-MATCHED LATENT FLOW

The emergent chemical semantics reveal that atom-wise latent representations from different elemental families tend to occupy largely non-overlapping regions of the VAE latent space. This observation suggests a natural route to controllable crystal generation: if sampled latent representations can be associated with a particular family-specific region, they may preferentially decode into atoms from the corresponding elemental family. To turn this observation into a generative mechanism, we first examine how source representations are associated with target families.

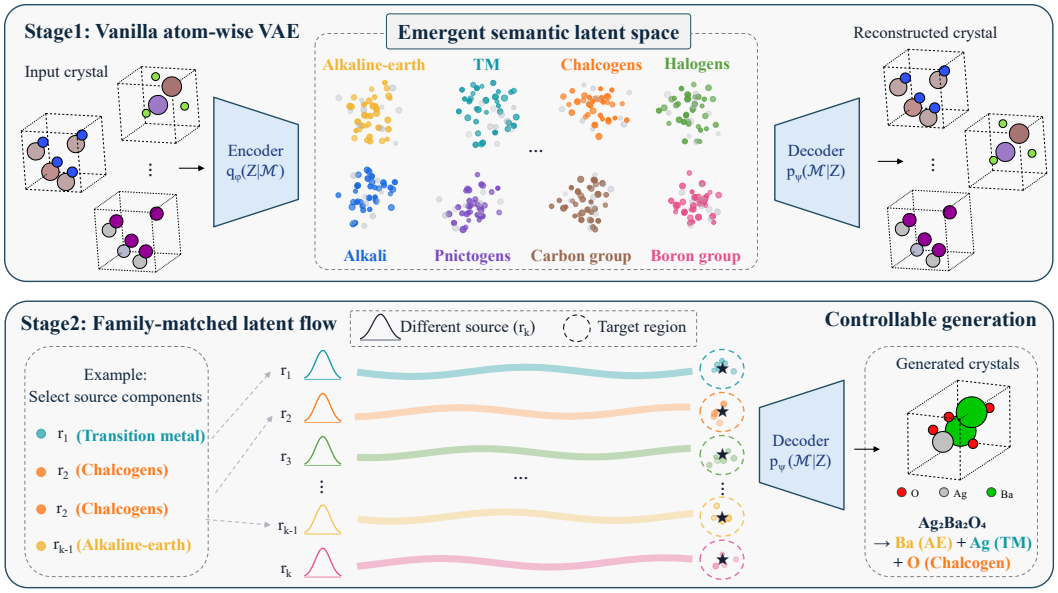


Figure 2: **Overview of FamilySpace.** *Vanilla atom-wise VAE:* Reconstruction and KL regularization yield atom-wise representations with emergent family organization. *Family-matched latent flow:* After the VAE is trained, source components are paired with posterior targets from the same family to train a shared flow model. *Controllable generation:* a family template selects the source component for each atomic representation. The shared flow transports the latent representations, and the frozen VAE decoder generates atomic species and crystal geometry.

Let $p_1^{\text{atom}}(\mathbf{z})$ denote the aggregated VAE posterior over atom-wise latent representations. Using family labels, it can be decomposed as

$$p_1^{\text{atom}}(\mathbf{z}) = \sum_{k \in \mathcal{K}_\mathcal{D}} \pi_k p_1^{\text{atom}}(\mathbf{z} \mid k), \tag{5}$$

where $\mathcal{K}_\mathcal{D}$ is the set of families observed in the training data, π_k is the corresponding weight, and $p_1^{\text{atom}}(\mathbf{z} \mid k)$ is the family-conditional aggregated posterior.

In standard latent flow training, source representations are sampled from the same isotropic Gaussian regardless of family:

$$p_0^{\text{atom}}(\mathbf{z}_0 \mid k) = p_0^{\text{atom}}(\mathbf{z}_0) = \mathcal{N}(\mathbf{z}_0; \mathbf{0}, \mathbf{I}), k \in \mathcal{K}_\mathcal{D}. \tag{6}$$

Thus, the source distribution of standard latent flow has no explicit correspondence with the family-specific regions of the VAE posterior, making this emergent semantics difficult to exploit directly for composition control.

To bridge this gap, FamilySpace introduces two changes to latent flow training: a family-indexed multi-source distribution and family-matched source–target coupling. Figure 2 illustrates the overall framework. FamilySpace retains the pretrained VAE, the flow backbone, and the standard flow matching objective, while modifying only these two components. We describe each below.

Family-indexed multi-source distribution. FamilySpace replaces the Gaussian source distribution with a family-indexed multi-source distribution. A deterministic mapping $g(a_i)$ assigns each atomic species to one of $K = 10$ families, as specified in Appendix A.1. For each observed family k, we select a representative posterior mean $\mathbf{m}_k$, referred to as the family medoid, and estimate its local variability from nearby same-family posterior means. Perturbations around $\mathbf{m}_k$ are scaled according to this local variability and bounded using the distances between family medoids. This construction defines a local source component r_k for each family, with the components constrained to have non-overlapping supports when the medoids are distinct. Details of the construction and sampling procedure are provided in Appendix B.1.

The resulting atom-level source distribution during training is

$$p_0^{\text{atom}}(\mathbf{z}) = \sum_{k \in \mathcal{K}_\mathcal{D}} \pi_k r_k(\mathbf{z}), \tag{7}$$

where each component r_k is indexed by an elemental family and π_k is the weight for each component. This multi-source distribution replaces only the source distribution used by the latent flow; the Gaussian prior used for VAE KL regularization remains unchanged.

Family-matched source-target coupling. A family-indexed source alone does not determine which family-specific region of the VAE posterior each source component should be transported toward. We therefore use family labels to construct family-matched source–target pairs during flow training. For a training crystal $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$ with N atoms, we get a target latent sequence from the frozen VAE posterior and select a source component for each atom according to its family:

$$\mathbf{Z}^{(1)} \sim q_\phi(\mathbf{Z} \mid \mathcal{M}), \qquad \mathbf{z}_i^{(0)} \sim r_{g(a_i)}, \qquad i = 1, \ldots, N. \tag{8}$$

Source construction uses posterior means, whereas flow training uses posterior samples as targets. Each source component is therefore paired only with targets from the same family:

$$r_k(\mathbf{z}_0) \longleftrightarrow p_1^{\text{atom}}(\mathbf{z}_1 \mid k), \qquad k \in \mathcal{K}_\mathcal{D}. \tag{9}$$

All target representations from a crystal remain together as a complete latent sequence. A single shared flow is trained on the paired source and target sequences using the linear path and standard flow matching objective defined in Section 3. Family labels determine the source components and training pairs, but are not provided as additional conditioning inputs to the flow network. Method details are provided in Appendix B. Training details are provided in Appendix B.2.

4.3 Controllable generation

At inference, a requested family template represented as

$$\mathbf{f} = (f_1, \ldots, f_N), \qquad f_i \in \mathcal{K}_\mathcal{D}, \tag{10}$$

specifies the desired family of each atomic representation. The number of representations assigned to each family determines the requested family counts and ratios. Given the template, source representations are sampled independently from the selected components:

$$p_0\left(\mathbf{Z}^{(0)} \mid \mathbf{f}\right) = \prod_{i=1}^{N} r_{f_i}\left(\mathbf{z}_i^{(0)}\right). \tag{11}$$

Thus, the template determines the family composition of the source sequence at inference. The shared latent flow transports the complete latent sequence, and the frozen VAE decoder generates atomic species, fractional coordinates, and lattice parameters.

Control is implemented through source selection. Each selected source component has been paired during training with targets from the requested family. The template specifies family composition, while exact element identities and crystal geometry remain generated by the model. Changing the template therefore directs generation toward a different family composition without further model training. We evaluate how accurately the decoded compositions follow the requested templates in Section 5.3. Sampling details are provided in Appendix B.3.

5 Experiments

5.1 Experimental setup

Our experiments evaluate the generation performance of FamilySpace, and the accuracy of family-composition control through source selection. We further evaluate whether chemically motivated family templates can support application-oriented exploration.

Datasets. We evaluate our method on the MP-20 dataset, which contains crystal structures with at most 20 atoms per primitive cell (Xie et al., 2022), and the MPTS-52 dataset (Baird et al., 2024),

Table 1: Crystal generation results on MP-20. The best results are highlighted in bold, while the second-best results are underlined.

Method	Valid (%) ↑	Uniqueness / novelty (%) ↑			Stability (%) ↑		Metastability (%) ↑	
		U	N	UN	S	SUN	MS	MSUN
DiffCSP (Jiao et al., 2023)	82.49	96.54	71.10	70.79	5.76	1.55	43.56	18.66
FlowMM (Miller et al., 2024)	82.14	97.86	72.34	72.20	4.91	1.40	39.43	15.80
MatterGen (Zeni et al., 2025)	83.58	**97.92**	74.96	**74.68**	4.96	1.80	44.74	23.23
ADiT (Joshi et al., 2025)	90.24	83.90	31.72	30.81	11.03	1.68	77.91	13.39
SymmCD (Levy et al., 2025)	84.60	97.87	73.19	72.79	4.84	1.53	34.09	10.90
SGEquiDiff (Chang et al., 2025)	85.42	96.56	63.68	63.36	6.52	1.46	44.77	12.36
CrystalDiT (Yi et al., 2026)	84.30	89.97	38.60	37.82	11.64	2.65	75.69	18.28
FamilySpace (Training templates)	89.48	97.14	69.63	69.53	6.26	1.38	42.99	15.31
FamilySpace (Alkali–Halogen 1:1)	**99.85**	47.74	58.40	45.25	**13.56**	0.70	**90.23**	36.12
FamilySpace (RE–Halogen 1:3)	91.22	64.53	**92.58**	63.00	7.87	**4.97**	70.36	**41.50**

Table 2: Crystal generation results on MPTS-52. The best results are highlighted in bold, while the second-best results are underlined.

Method	Valid (%) ↑	Uniqueness / novelty (%) ↑			Stability (%) ↑		Metastability (%) ↑	
		U	N	UN	S	SUN	MS	MSUN
ADiT (Joshi et al., 2025)	83.82	93.42	55.64	55.15	9.77	2.32	59.21	18.33
SGEquiDiff (Chang et al., 2025)	80.59	**97.69**	89.07	**88.18**	4.31	2.39	26.13	16.05
FamilySpace (Training templates)	78.59	96.75	84.47	83.73	5.32	2.11	31.60	17.27
FamilySpace (Alkali–Halogen 1:1)	97.32	24.56	39.65	24.14	**18.75**	0.63	**93.89**	21.10
FamilySpace (RE–Halogen 1:3)	**98.11**	69.56	**95.22**	69.28	8.97	**5.10**	74.55	**46.24**

which consists of crystal structures with up to 52 atoms per unit cell. For both datasets, we follow the training, validation, and test splits adopted in prior work (Miller et al., 2024).

Baselines. We compare FamilySpace with DiffCSP (Jiao et al., 2023), FlowMM (Miller et al., 2024), MatterGen (Zeni et al., 2025), ADiT (Joshi et al., 2025), SymmCD (Levy et al., 2025), SGEquiDiff (Chang et al., 2025), and CrystalDiT (Yi et al., 2026) on MP-20. On MPTS-52, we compare with ADiT and SGEquiDiff.

Implementation. Since FamilySpace is designed for controllable generation, it requires a family template specifying the requested family composition of the generated crystals. We therefore evaluate FamilySpace under different templates. For de novo generation, we consider three representative settings. The first is *training-template sampling*, under which 10,000 samples are generated by drawing family compositions from the empirical training distribution. The other two are the *Alkali–Halogen (1:1)* and *RE–Halogen (1:3)* templates[1], which specify the corresponding family ratios and are selected as chemically motivated targets for stability-oriented exploration. The template-selection procedure is in Appendix D.2. Generation experiments use a separately trained Transformer VAE for each dataset and a single flow network shared across families within each dataset. The VAE remains frozen during flow training and generation. Architecture and training settings are provided in Appendices A.2 and B.2, and sampling details are given in Appendix B.3.

Evaluations. We evaluate generated structures in terms of validity, uniqueness and novelty, and thermodynamic stability. A crystal is considered valid only when it satisfies both compositional and structural validity criteria. Uniqueness (U) measures whether a generated crystal is distinct from other generated samples, while novelty (N) measures whether it is absent from the reference structure database; we further report their intersection $\mathrm{UN} = \mathrm{U} \cap \mathrm{N}$. Thermodynamic stability is evaluated after MatterSim relaxation (`MatterSim-v1.0.0-1M`) (Yang et al., 2024; Zeni et al., 2025) using the energy above the convex hull, E_{hull}. We define stable samples as $\mathrm{S} : E_{\mathrm{hull}} \leq 0\ \mathrm{eV/atom}$ and stable or metastable samples as $\mathrm{MS} : E_{\mathrm{hull}} \leq 0.1\ \mathrm{eV/atom}$. Accordingly, we

[1]For family combinations, AE denotes Alkaline earth, TM Transition metal, RE Rare earth, Ch Chalcogens, and Pn Pnictogens, following the family mapping in Appendix A.1.

Table 3: Single-family control on MP-20. *Target-family fraction* measures the fraction of decoded atoms belonging to the requested family. *Family-count accuracy* measures the percentage of samples whose decoded family-count vector exactly matches the requested one. Formal metric definitions are provided in Appendix C.4.

Target family	Target-family fraction (%)	Family-count accuracy (%)
Alkali	99.44	96.90
Alkaline earth	99.90	99.30
Boron group	99.97	99.80
Carbon group	99.93	99.70
Pnictogens	99.98	99.70
Chalcogens	99.92	99.50
Halogen	99.87	99.20
Noble gas	91.78	68.80
Transition metal	99.90	99.50
Rare earth	97.92	89.60
All families	98.86	95.20

report $\mathrm{SUN} = \mathrm{S} \cap \mathrm{UN}$ and $\mathrm{MSUN} = \mathrm{MS} \cap \mathrm{UN}$ to quantify candidates that are simultaneously thermodynamically favorable, unique, and novel. More details are provided in Appendix C.

5.2 De novo generation

Table 1 compares FamilySpace with baseline methods on the MP-20 dataset. Under the training-template sampling, FamilySpace achieves a substantially higher UN rate than ADiT, resulting in a higher MSUN rate. More importantly, the controllable generation capability of FamilySpace enables direct stability-oriented exploration of latent regions by simply selecting appropriate family templates, without requiring additional optimization techniques such as reinforcement learning. With the Alkali–Halogen (1:1) template, the S and MS rates increase to 13.56% and 90.23%, respectively, exceeding all evaluated baselines. The RE–Halogen (1:3) template further achieves SUN and MSUN rates of 4.97% and 41.50%, respectively, again exceeding the corresponding baseline rates. Table 2 compares FamilySpace with baseline methods on the MPTS-52 dataset. The results show trends consistent with those observed on MP-20. These results demonstrate that FamilySpace can achieve strong generation performance.

5.3 Controllable generation

In this section, we evaluate whether selecting source components produces the requested family composition after latent transport and decoding. All control metric definitions are provided in Appendix C.4.

Single-family control. We first evaluate each source component individually. For a requested family, all source representations are sampled from its component, transported by the shared flow, and decoded into a crystal. We generate 1,000 samples for each of the ten families. As shown in Table 3, the average *Target-family fraction* is 98.86%, and the overall *Family-count accuracy* is 95.20%. For nine families, the Target-family fraction exceeds 97.9%. The noble-gas family has the lowest Family-count accuracy (68.80%) and contains only 38 training atoms. Excluding noble gases, Family-count accuracy is 98.13%. These results show a strong empirical correspondence between source-component selection and decoded family composition.

Multi-family ratio control. We next combine source components to generate crystals with prescribed family ratios. We evaluate five family combinations containing two to four families, with 10,000 samples per combination. For example, 1:1:2 and 2:2:4 represent the same reduced ratio. As shown in Table 4, FamilySpace achieves 99.992% Family-set accuracy and 99.968% Family-ratio accuracy across all 50,000 samples. These results show that source components can be combined to control the requested family composition with high accuracy. In addition, family-level control does not fix the exact elements. The corresponding diversity results are reported in Appendix D.1.

Table 4: Multi-family ratio control on MP-20. *Family-set accuracy* measures whether the generated crystal contains exactly the requested families, while *Family-ratio accuracy* additionally requires the reduced family ratio to match the request. Formal metric definitions are provided in Appendix C.4.

Family combination	Tested ratios	Family-set accuracy (%)	Family-ratio accuracy (%)
TM–Ch	1:1, 1:2, 2:3, 3:4	99.99	99.94
Alkali–TM–Ch	1:1:2, 1:2:4, 1:1:3, 2:1:3	99.99	99.95
AE–TM–Ch	1:1:3, 1:2:4, 1:1:2	100.00	99.98
RE–TM–Ch	1:1:3, 1:1:4, 2:1:4	99.99	99.98
Alkali–TM–Pn–Ch	1:1:1:4, 1:1:2:7, 2:1:1:4	99.99	99.99
All 50,000 samples		99.992	99.968

Application-oriented exploration. Beyond reproducing prescribed family compositions, FamilySpace allows generation to be directed toward chemically motivated regions of materials space. We examine whether this capability enables access to compositions associated with reported functional materials. We consider the five family combinations in Table 4, generate 10,000 structures for each combination, and retain candidates satisfying the MSUN criterion. Detailed screening results are reported in Appendix D.3.

We conduct a composition-level literature comparison to identify generated candidates whose compositions have been investigated for specific applications in recent studies. Representative examples span several family templates. The Alkali–TM–Ch template produces $\mathrm{NaCrO_2}$, which has been investigated as a sodium-ion battery cathode, with improved rate capability and cycling stability achieved through nitrogen-doped carbon coating (Cai et al., 2025). The AE–TM–Ch template yields $\mathrm{SrTiO_3}$, whose reported single-crystal form has been investigated for photocatalytic overall water splitting (Luo et al., 2025a), and $\mathrm{BaZrO_3}$, a perovskite host relevant to proton-conducting ceramic electrolytes (Tang et al., 2025). The RE–TM–Ch template yields $\mathrm{LaMnO_3}$, a composition investigated for oxygen-evolution electrocatalysis in the form of unconventional cubic nanosheets (Dai et al., 2025). The Alkali–TM–Pn–Ch template produces $\mathrm{Li_2MnFe(PO_4)_2}$, which is stoichiometrically equivalent to $\mathrm{LiMn_{0.5}Fe_{0.5}PO_4}$, a composition studied as a lithium-ion battery cathode with engineered carbon coatings (Hou et al., 2025). These examples show that family-controlled generation can access compositions associated with energy storage, energy conversion, and catalysis. Visualizations of representative generated structures are provided in Appendix D.3.

This analysis provides a preliminary assessment of the application relevance of the generated compositions. Composition-level correspondence does not establish equivalence to the reported crystal phases or functional performance, which may also depend on doping, coatings, surface structure, and morphology. DFT calculations and experimental validation are beyond the scope of this study. Further comparison with reference crystal structures, property calculations, and experimental characterization would be needed to assess the practical potential of the generated candidates.

6 Conclusion

In this work, we investigated the geometric organization of atom-wise latent representations learned by a vanilla crystal VAE and showed that they exhibit emergent chemical semantics: representations from the same elemental family form coherent neighborhoods, while different families occupy largely separated regions of the latent space. Building on this observation, we introduced FamilySpace, which converts this latent organization into a family-indexed multi-source distribution and uses family-matched source–target coupling to enable controllable latent flow generation. Experiments on MP-20 and MPTS-52 demonstrate that FamilySpace maintains competitive crystal-generation performance while accurately following prescribed elemental-family compositions. Moreover, chemically motivated family templates can direct generation toward regions with higher fractions of candidates predicted to be stable or metastable, without additional optimization such as reinforcement learning. These results show that the internal organization of learned crystal latent representations can provide a useful control mechanism for generative modeling and support targeted exploration of materials chemical design space.

REFERENCES

Sterling G Baird, Hasan M Sayeed, Joseph Montoya, and Taylor D Sparks. matbench-genmetrics: A python library for benchmarking crystal structure generative models using time-based splits of materials project structures. *Journal of Open Source Software*, 9(97):5618, 2024.

Tian Bian, Yifan Niu, Heng Chang, Divin Yan, Junzhou Huang, Yu Rong, Tingyang Xu, Jia Li, and Hong Cheng. Hierarchical graph latent diffusion model for conditional molecule generation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 130–140, 2024.

Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.

Jingyao Cai, Yan Wei, Yanbin Zhu, Liyuan Tian, Jiandong Zhang, Yan Shen, and Mingkui Wang. Nitrogen-doped carbon-coating enables high-rate capability and long-cycle stability of nacro2 cathode for sodium-ion battery. *Applied Surface Science*, 683:161804, 2025.

Rees Chang, Angela Pak, Alex Guerra, Ni Zhan, Nick Richardson, Elif Ertekin, and Ryan Adams. Space group equivariant crystal diffusion. *Advances in Neural Information Processing Systems*, 38:72772–72805, 2025.

Mouyang Cheng, Chu-Liang Fu, Ryotaro Okabe, Abhijatmedhi Chotrattanapituk, Artittaya Boonkird, Nguyen Tuan Hung, and Mingda Li. Artificial intelligence-driven approaches for materials design and discovery. *Nature Materials*, pp. 1–17, 2026.

François R J Cornet, Federico Bergamin, Arghya Bhowmik, Juan Maria Garcia-Lastra, Jes Frellsen, and Mikkel N. Schmidt. Kinetic langevin diffusion for crystalline materials generation. In *Forty-second International Conference on Machine Learning*, 2025.

Jiao Dai, Mingjie Wang, Haoran Tian, Wenhui Fan, Kaisi Liu, Weilin Xu, Huiyu Jiang, Huanyu Jin, and Jun Wan. Microwave shock-driven thermal engineering of unconventional cubic 2d lamno3 for efficient oxygen evolution. *Journal of Materials Chemistry A*, 13(37):31002–31012, 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 *International Conference on Learning Representations*, volume 2025, pp. 14080–14110, 2025.

Yuhui Ding and Thomas Hofmann. Scalable non-equivariant 3d molecule generation via rotational alignment. In *Forty-second International Conference on Machine Learning*, 2025.

Magdolna Hargittai. Molecular structure of metal halides. *Chemical reviews*, 100(6):2233–2302, 2000.

Philipp Höllmer and Stefano Martiniani. Open materials generation with inference-time reinforcement learning. In *Forty-third International Conference on Machine Learning*, 2026.

Philipp Höllmer, Thomas Egg, Maya Martirossyan, Eric Fuemmeler, Zeren Shui, Amit Gupta, Pawan Prakash, Adrian Roitberg, Mingjie Liu, George Karypis, Mark Transtrum, Richard Hennig, Ellad B. Tadmor, and Stefano Martiniani. Open materials generation with stochastic interpolants. In *Forty-second International Conference on Machine Learning*, 2025.

Ying Hou, Yuxing Xu, Shiyang Jiang, Yuncheng Zhou, Jiechen Song, Chen Wu, and Qiangqiang Tan. Precursor-driven carbon layer construction enables high-performance limn0. 5fe0. 5po4 cathodes. *Electrochimica Acta*, 533:146433, 2025.

Anubhav Jain, Geoffroy Hautier, Charles J Moore, Shyue Ping Ong, Christopher C Fischer, Tim Mueller, Kristin A Persson, and Gerbrand Ceder. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science*, 50(8):2295–2310, 2011.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL `https://github.com/jiaor17/DiffCSP`.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary Ward Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *Forty-second International Conference on Machine Learning*, 2025.

Nikita Kazeev, Wei Nong, Ignat Romanov, Ruiming Zhu, Andrey E Ustyuzhanin, Shuya Yamazaki, and Kedar Hippalgaonkar. Wyckoff transformer: Generation of symmetric crystals. In *Forty-second International Conference on Machine Learning*, 2025.

Filip Ekström Kelvinius, Oskar B. Andersson, Abhijith S Parackal, Dong Qian, Rickard Armiento, and Fredrik Lindsten. Wyckoffdiff – a generative diffusion model for crystal symmetry. In *Forty-second International Conference on Machine Learning*, 2025.

Subhojyoti Khastagir, Kishalay Das, Pawan Goyal, Seung-Cheol Lee, Satadeep Bhattacharjee, and Niloy Ganguly. Crysldm: Latent diffusion model for crystal material generation. *AI for Accelerated Materials Design-ICLR*, 2025, 2025a.

Subhojyoti Khastagir, Kishalay Das, Pawan Goyal, Seung-Cheol Lee, Satadeep Bhattacharjee, et al. Llm meets diffusion: a hybrid framework for crystal material generation. *Advances in Neural Information Processing Systems*, 38:169190–169215, 2025b.

PN Kumta and SH Risbud. Rare-earth chalcogenides—an emerging class of optical materials. *Journal of materials science*, 29(5):1135–1158, 1994.

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 *International Conference on Learning Representations*, volume 2025, pp. 23166–23189, 2025.

Zhenzhu Li and Aron Walsh. Platonic representation of foundation machine learning interatomic potentials. *Nature Machine Intelligence*, pp. 1–11, 2026.

Yaling Luo, Chenwei Ni, Jianfeng Zhao, Junhao Cui, Jie Zhang, Qianhong Zhu, Fengtao Fan, and Can Li. Intrafacet charge separation toward efficient overall water splitting on srtio3 single crystal photocatalysts. *The Journal of Physical Chemistry Letters*, 16(12):2981–2986, 2025a.

Yanchen Luo, Zhiyuan Liu, Yi Zhao, Sihang Li, Hengxing Cai, Kenji Kawaguchi, Tat-Seng Chua, Yang Zhang, and Xiang Wang. Towards unified and lossless latent space for 3d molecular latent diffusion modeling. *Advances in Neural Information Processing Systems*, 38:30718–30748, 2025b.

Youzhi Luo, Chengkai Liu, and Shuiwang Ji. Towards symmetry-aware generation of periodic materials. *Advances in Neural Information Processing Systems*, 36:53308–53329, 2023.

Wilhelm F Maier. Early years of high-throughput experimentation and combinatorial approaches in catalysis and materials science. *ACS combinatorial science*, 21(6):437–444, 2019.

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M Wood. FlowMM: generating materials with riemannian flow matching. In *Forty-first International Conference on Machine Learning*, 2024.

Kwasi Mitchell and James A Ibers. Rare-earth transition-metal chalcogenides. *Chemical reviews*, 102(6):1929–1952, 2002.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

Hyunsoo Park and Aron Walsh. Guiding generative models to uncover diverse and novel crystals via reinforcement learning. *Nature Machine Intelligence*, pp. 1–13, 2026.

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

Radislav Potyrailo, Krishna Rajan, Klaus Stoewe, Ichiro Takeuchi, Bret Chisholm, and Hubert Lam. Combinatorial and high-throughput screening of materials libraries: review of state of the art. *ACS combinatorial science*, 13(6):579–633, 2011.

Kiyoung Seong, Sungsoo Ahn, Sehui Han, and Changyoung Park. Multimodal crystal flow: Any-to-any modality generation for unified crystal modeling. In *Forty-third International Conference on Machine Learning*, 2026.

Dinker B Sirdeshmukh, Lalitha Sirdeshmukh, and KG Subhadra. *Alkali Halides: A Handbook of physical properties*, volume 49. Springer Science & Business Media, 2001.

Boris N Slautin, Utkarsh Pratiush, Doru C Lupascu, Maxim A Ziatdinov, and Sergei V Kalinin. Integrating predictive and generative capabilities by latent space design via the dkl-vae model. *npj Computational Materials*, 12(1):231, 2026.

Anuroop Sriram, Benjamin Miller, Ricky TQ Chen, and Brandon Wood. Flowllm: Flow matching for material generation with large language models as base distributions. In *Advances in Neural Information Processing Systems*, volume 37, pp. 46025–46046, 2024.

Huihong Tang, Chu Chen, Zihao Chen, Zhishan Li, Meiting Guo, Zehua Wang, Zongping Shao, Zhenhang Liu, Yu Shen, Zhongwei Yue, et al. Phase and microstructure stability of ba-containing proton conducting electrolyte materials under reducing environment of protonic ceramic cells-baceo3 and bazro3. *Ceramics International*, 2025.

Conghao Wang, Hiok Hian Ong, Shunsuke Chiba, and Jagath C Rajapakse. Gldm: hit molecule generation with constrained graph latent diffusion model. *Briefings in bioinformatics*, 25(3): bbae142, 2024.

Zhenzhong Wang, Haowei Hua, Wanyu Lin, Ming Yang, and Kay Chen Tan. Crystalline material discovery in the era of artificial intelligence. *ACM Computing Surveys*, 58(10):1–37, 2026.

Tian Xie, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi S. Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations*, 2022.

Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, Matthew Horton, Robert Pinsler, Andrew Fowler, Daniel Zügner, Tian Xie, Jake Smith, Lixin Sun, Qian Wang, Lingyu Kong, Chang Liu, Hongxia Hao, and Ziheng Lu. Mattersim: A deep learning atomistic model across elements, temperatures and pressures, 2024.

Xiaohan Yi, Guikun Xu, Zhong Zhang, Liu Liu, Yatao Bian, Xi Xiao, and Peilin Zhao. Crystaldit: Simple diffusion transformers for crystal generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1462–1470, 2026.

Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, 2025.

Qunhao Zhang, Jun Xiao, Dongjiang Niu, Zhixin Zhang, Shanyang Ding, and Zhen Li. Geometry-complete latent diffusion model for 3d molecule generation. *Bioinformatics*, 41(8):btaf426, 2025.

Yong Zhao, Mohammed Al-Fahdi, Ming Hu, Edirisuriya MD Siriwardane, Yuqi Song, Alireza Nasiri, and Jianjun Hu. High-throughput discovery of novel cubic crystal materials using deep generative neural networks. *Advanced Science*, 8(20):2100566, 2021.

Matvei Zinkevich. Thermodynamics of rare earth sesquioxides. *Progress in Materials Science*, 52 (4):597–647, 2007.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language polishing, experimental code development, and literature search. All AI-assisted text was reviewed by the authors. All AI-assisted code was tested and verified to work as expected. Relevant literature identified through AI-assisted searches was verified against the original sources.

APPENDIX

A VAE DETAILS

A.1 PERIODIC-TABLE FAMILY MAPPING

Let $\mathcal{K} = \{0, \dots, K-1\}$, with $K = 10$, denote the fixed family vocabulary. Table 5 specifies the lookup $g : a \mapsto k$ used throughout the paper. Family labels are introduced *after* VAE training for representation analysis, source construction, and source–target coupling. At inference, a requested family template selects from $\mathcal{K}_{\mathcal{D}} \subseteq \mathcal{K}$, the families observed in the training data.

Table 5: The deterministic family lookup used by FamilySpace. The category names denote the implemented partition: hydrogen is assigned to Alkali, and Rare earth includes both lanthanides and actinides. These assignments are modeling conventions.

ID	Family	Elements
0	Alkali	H, Li, Na, K, Rb, Cs, Fr
1	Alkaline earth (AE)	Be, Mg, Ca, Sr, Ba, Ra
2	Boron group	B, Al, Ga, In, Tl, Nh
3	Carbon group	C, Si, Ge, Sn, Pb, Fl
4	Pnictogens (Pn)	N, P, As, Sb, Bi, Mc
5	Chalcogens (Ch)	O, S, Se, Te, Po, Lv
6	Halogen	F, Cl, Br, I, At, Ts
7	Noble gas	He, Ne, Ar, Kr, Xe, Rn, Og
8	Transition metal (TM)	Groups 3–12, excluding lanthanides and actinides
9	Rare earth (RE)	Lanthanides and actinides

A.2 VANILLA ATOM-WISE VAE

Both datasets use an atom-wise Transformer VAE with an eight-dimensional latent for each atom. The encoder and decoder each have eight layers, hidden width 512, eight attention heads, and feed-forward width 2,048. The encoder predicts a diagonal Gaussian posterior, and the decoder reconstructs atomic species, fractional coordinates, cell lengths, and cell angles. The architecture is shared across datasets, but the model parameters are learned separately using dataset-specific reconstruction objectives.

MP-20. The VAE uses

$$\mathcal{L}_{\mathrm{VAE}}^{\mathrm{MP20}} = 2\mathcal{L}_A + 5\mathcal{L}_F^{\mathrm{PBC}} + \mathcal{L}_l + 10\mathcal{L}_\alpha + 10^{-5}\mathcal{L}_{\mathrm{KL}}. \tag{12}$$

Atomic species use cross-entropy. The coordinate residual is the componentwise minimum-image displacement in fractional space:

$$\Delta\mathbf{f}_i = \mathrm{remainder}\Big(\widehat{\mathbf{f}}_i - \mathbf{f}_i + \tfrac{1}{2}, 1\Big) - \tfrac{1}{2}, \qquad \ell_F^{\mathrm{PBC}}(i) = \tfrac{1}{3}\|\Delta\mathbf{f}_i\|_2^2. \tag{13}$$

The factor $1/3$ follows the implementation's mean over coordinate components. This residual does not weight fractional displacements by the cell metric. The VAE is trained for 2,000 epochs.

MPTS-52. The MPTS-52 VAE uses ordinary squared error on fractional coordinates:

$$\ell_F^{\mathrm{ordinary}}(i) = \tfrac{1}{3}\|\widehat{\mathbf{f}}_i - \mathbf{f}_i\|_2^2, \qquad \mathcal{L}_{\mathrm{VAE}}^{\mathrm{MPTS52}} = \mathcal{L}_A + 10\mathcal{L}_F^{\mathrm{ordinary}} + \mathcal{L}_l + 10\mathcal{L}_\alpha + 10^{-5}\mathcal{L}_{\mathrm{KL}}. \tag{14}$$

The training run was configured for 2,000 epochs. All MPTS-52 generation results use this same frozen VAE.

Optimization and augmentation. Coordinate and atom-type losses are averaged over atoms in the batch. Length and angle losses first average over their three components and then over structures. The KL term sums over the eight latent dimensions and averages over atoms.

Training applies a random global translation and a common rotation of the Cartesian positions and cell. Independently, 10% of atomic-species inputs are masked and 10% of Cartesian positions receive Gaussian noise with standard deviation 0.1 Å per component. Fractional coordinates are recomputed and wrapped into the unit cell. Targets retain the global augmentation but exclude the masking and added noise. Both runs use AdamW with learning rate 10^{-4}, zero weight decay, $(\beta_1, \beta_2) = (0.9, 0.999)$, and $\epsilon = 10^{-8}$. There is no learning-rate schedule.

Table 6: VAE architecture and optimization settings.

Setting	MP-20	MPTS-52
Encoder / decoder layers	8 / 8	8 / 8
Hidden / latent width	512 / 8	512 / 8
Attention heads / feed-forward width	8 / 2048	8 / 2048
Activation / normalization	GELU / pre-norm	GELU / pre-norm
Dropout	0	0
Optimizer / learning rate	AdamW / 10^{-4}	AdamW / 10^{-4}
Weight decay / learning-rate scheduler	0 / none	0 / none
Batch size / Configured epochs	128 / 2000	256 / 2000
Gradient-norm clipping	10	10

A.3 Representation extraction and visualization

Representation extraction. We extract the posterior mean of every atom in the test split, obtaining 93,158 representations from 9,046 crystals for MP-20 and 202,522 representations from 8,096 crystals for MPTS-52. Each dataset is encoded using its corresponding frozen VAE. The posterior means, rather than stochastic posterior samples, provide the deterministic atom-wise representations used for visualization. For both datasets, t-SNE is applied with perplexity 40 and 1500 iterations.

We additionally visualize the posterior representations extracted from the complete training splits using the same procedure. The MP-20 training split contains 282,847 atoms from 27,136 crystals, while the MPTS-52 training split contains 501,334 atoms from 27,380 crystals. As shown in Figure 3, the family-level organization observed in the test representations is also clearly present in the training representations.

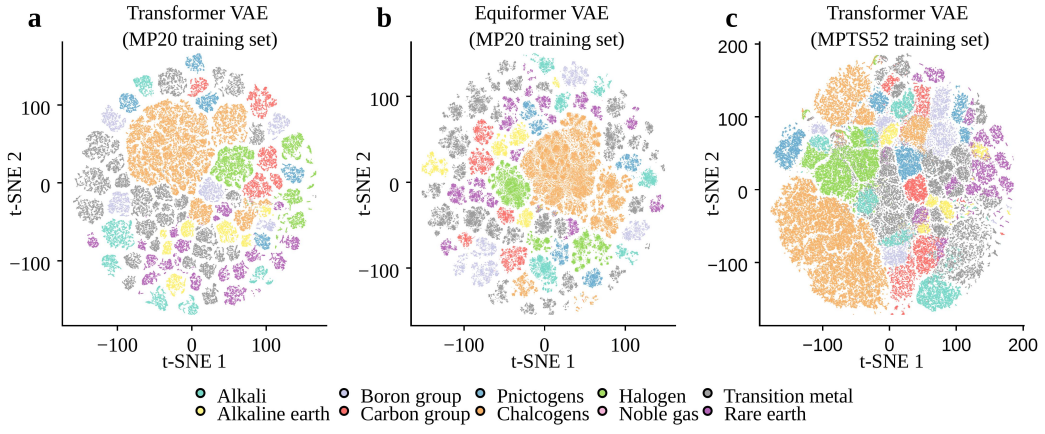


Figure 3: **Emergent chemical semantics in atom-wise latent representations on the training sets.** We visualize the posterior-mean atom-wise latent representations extracted from the complete training splits using t-SNE. (a) Transformer VAE on MP-20. (b) Equiformer VAE on MP-20. (c) Transformer VAE on MPTS-52. All representations are colored by periodic-table family.

Table 7: Quantitative family-neighborhood analysis of Transformer VAE posterior means on the MP-20 and MPTS-52 test sets. N denotes the number of test atoms in each family.

Family	MP-20			MPTS-52		
	N	Family 1-NN (%)	Family 5-NN (%)	N	Family 1-NN (%)	Family 5-NN (%)
Alkali	6,647	99.8195	98.9619	20,206	98.0006	90.2108
Alkaline earth	4,322	99.6992	98.9125	8,382	97.3157	87.5447
Boron group	5,724	99.9651	99.8078	11,075	96.2799	87.0429
Carbon group	5,754	99.9305	99.5829	12,791	97.9595	90.8060
Pnictogens	4,848	99.9587	99.6700	15,444	98.9964	94.3603
Chalcogens	23,824	99.9706	99.8531	79,150	99.7688	98.8932
Halogen	8,598	99.4766	96.9877	14,282	98.1725	91.3387
Noble gas	11	36.3636	0.0000	81	40.7407	0.0000
Transition metal	23,997	99.6666	98.4665	26,765	95.7183	83.6279
Rare earth	9,433	99.6396	98.1024	14,346	96.8144	86.4213
All atoms	93,158	99.78	98.91	202,522	98.25	92.58

Because the ADiT VAE architecture is also applicable to molecular generation, we further visualize the latent space learned by a VAE trained on the QM9 molecular dataset. As shown in Figure 4, semantically meaningful organization also emerges in the latent space of the VAE trained on QM9. This observation suggests that emergent semantic organization may be a more general phenomenon in atom-wise latent spaces for atomic systems. In this work, we mainly focus on crystal generation; extending this investigation to molecules and other atomistic systems is an interesting direction for future work.

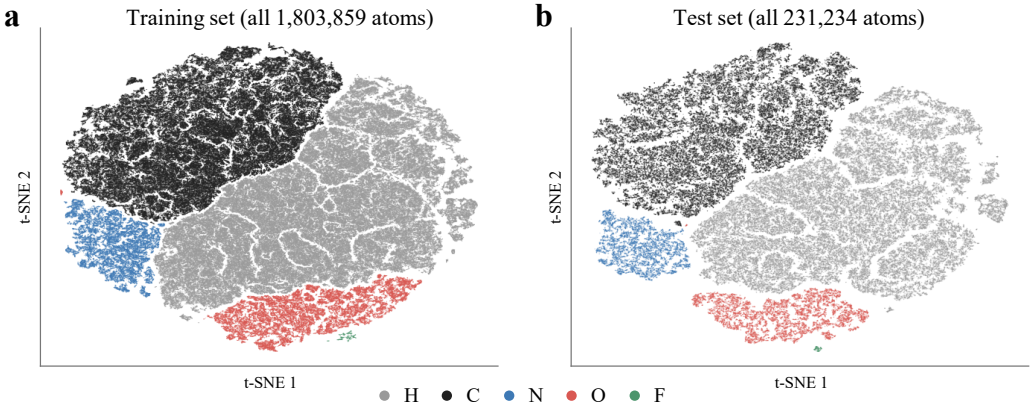


Figure 4: **Emergent chemical semantics in atom-wise latent representations on QM9.** We visualize the posterior-mean atom-wise latent representations extracted from the complete (a) training split and (b) test split, respectively.

Quantitative family-neighborhood analysis. To complement the qualitative results in Figure 1, we further quantify the local family-level organization of the Transformer VAE posterior means on MP-20 and MPTS-52. We evaluate the posterior mean representation using all atoms in the test split, including 93,158 atoms for MP-20 and 202,522 atoms for MPTS-52. Nearest neighbors are identified using exact Euclidean distance in the posterior space. We report two nearest-neighbor metrics. *Family 1-NN* measures the percentage of atoms whose nearest neighbor belongs to the same periodic-table family. *Family 5-NN* measures the percentage of atoms for which all five nearest neighbors belong to the same periodic-table family as the query atom. As shown in Table 7, the posterior representations exhibit strong local family-level organization on both datasets. On MP-20, all well-represented families achieve above 95% accuracy for both nearest-neighbor metrics. MPTS-52 shows slightly weaker but still strong local organization.

B FAMILYSPACE DETAILS

FamilySpace is constructed after the VAE has been trained and frozen. The VAE retains its standard Gaussian prior for KL regularization; the family-indexed multi-source distribution replaces only the source distribution of the subsequent latent flow. This section specifies source construction, training, and sampling.

B.1 SOURCE CONSTRUCTION

Our goal is to construct a local source distribution for each periodic-table family using the posterior representations of the frozen VAE. Each source component is specified by three quantities: a representative center $\mathbf{m}_k$, a scale $\mathbf{s}_k$, and a support radius ρ_k. We first select an observed posterior representation near the center of each family, then estimate the local variation around it using nearby same-family representations. We use these estimates to sample perturbations around the representative and bound their magnitude according to the distance to neighboring family centers. Together, these steps produce the family-indexed source components.

Family representative center estimation. Let $\boldsymbol{\mu}_i \in \mathbb{R}^d$ denote the posterior mean of training atom i, and let $\boldsymbol{\mu}_{\text{glob}}$ and $\boldsymbol{\sigma}_{\text{glob}}$ denote the mean and standard deviation over all training atoms. Let $\mathcal{I}_k = \{i : g(a_i) = k\}$ denote the indices of training atoms belonging to family k. We standardize each posterior representation as

$$\widetilde{\boldsymbol{\mu}}_i = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\text{glob}}) \oslash \boldsymbol{\sigma}_{\text{glob}}, \tag{15}$$

where $\oslash$ denotes coordinate-wise division. For each observed family k, we compute its centroid in the standardized latent space and select the representation nearest to this centroid:

$$\overline{\widetilde{\boldsymbol{\mu}}}_k = \frac{1}{|\mathcal{I}_k|} \sum_{i \in \mathcal{I}_k} \widetilde{\boldsymbol{\mu}}_i, \qquad i_k^* = \arg\min_{i \in \mathcal{I}_k} \left\| \widetilde{\boldsymbol{\mu}}_i - \overline{\widetilde{\boldsymbol{\mu}}}_k \right\|_2, \qquad \mathbf{m}_k = \boldsymbol{\mu}_{i_k^*}. \tag{16}$$

The term *family medoid* in this paper refers to this centroid-nearest representation $\mathbf{m}_k$.

Local scale estimation. For each family k, we select up to 1,000 same-family representations closest to its medoid, using Euclidean distance in the standardized latent space. We denote this neighborhood by $\mathcal{N}_k$ and use all available representations when the family contains fewer than 1,000 atoms. We then estimate the local scale along each coordinate j from the original, unstandardized latent representations:

$$s_{k,j} = \max\left\{ \text{Std}_{i \in \mathcal{N}_k}(\mu_{i,j}), 10^{-5} \right\}. \tag{17}$$

The lower bound prevents the scale from vanishing. The resulting vector $\mathbf{s}_k$ captures the local variation along each latent dimension and determines the relative magnitude of the sampling perturbations.

Bounded local sampling. Let r_k denote the source component for family k and ρ_k its support radius. For distinct family medoids, set

$$\rho_k = \beta \min_{\ell \in \mathcal{K}_{\mathcal{D}},\, \ell \neq k} \|\mathbf{m}_k - \mathbf{m}_\ell\|_2, \qquad \beta = 0.2. \tag{18}$$

A sample from r_k is generated as

$$\begin{aligned} &\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), && \boldsymbol{\delta}_k = \alpha \mathbf{s}_k \odot \boldsymbol{\epsilon}, \quad \alpha = 0.1, \\ &\mathbf{z}^{(0)} = \mathbf{m}_k + \boldsymbol{\delta}_k \min\left(1, \frac{\rho_k}{\|\boldsymbol{\delta}_k\|_2 + 10^{-12}}\right). \end{aligned} \tag{19}$$

This operation projects a locally scaled Gaussian perturbation into a bounded region centered at the family medoid. With $\beta = 0.2 < 1/2$, the resulting source components have pairwise disjoint supports when the family medoids are distinct.

B.2 FAMILY-MATCHED LATENT FLOW TRAINING

For each training crystal $\mathcal{M} = (\mathbf{A}, \mathbf{F}, \mathbf{L})$, we sample a target sequence $\mathbf{Z}^{(1)} \sim q_\phi(\mathbf{Z} \mid \mathcal{M})$ from the frozen VAE. Each target atom selects a source $\mathbf{z}_i^{(0)} \sim r_{g(a_i)}$. Consequently, the targets used to train the flow are posterior *samples*, whereas the family medoids and local scales in Appendix B.1 are estimated from posterior *means*.

We draw $t \sim \mathcal{U}(0.01, 0.99)$ and construct the linear path in Equation 2. Consistently with Section 3, the network predicts the clean endpoint, and the velocity is obtained from that prediction:

$$\widehat{\mathbf{Z}}^{(1)} = F_\theta(\mathbf{Z}^{(t)}, t), \qquad \mathbf{v}_\theta(\mathbf{Z}^{(t)}, t) = \frac{F_\theta(\mathbf{Z}^{(t)}, t) - \mathbf{Z}^{(t)}}{1 - t}. \tag{20}$$

The sole training objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{FM}} &= \mathbb{E}\left[\frac{1}{Nd} \sum_{i=1}^{N} \left\| \mathbf{v}_\theta(\mathbf{Z}^{(t)}, t)_i - (\mathbf{z}_i^{(1)} - \mathbf{z}_i^{(0)}) \right\|_2^2 \right] \\ &= \mathbb{E}\left[\frac{\|F_\theta(\mathbf{Z}^{(t)}, t) - \mathbf{Z}^{(1)}\|_F^2}{Nd(1-t)^2} \right]. \end{aligned} \tag{21}$$

The implementation averages over unpadded atoms and all $d = 8$ latent dimensions, followed by an average over crystals in the minibatch.

Following Joshi et al. (2025), we use a DiT-based flow network with 12 layers, a hidden dimension of 768, and 12 attention heads. An exponential moving average (EMA) of the flow parameters is updated after each optimization step and used for generation. The VAE remains frozen throughout flow training. Complete training and sampling settings are provided in Table 8.

Table 8: Latent-flow settings for both datasets. The frozen VAE and the family sources are specific to each dataset.

Setting	MP-20	MPTS-52
Source noise α / radius factor β	0.1 / 0.2	0.1 / 0.2
DiT layers / hidden width / attention heads	12 / 768 / 12	12 / 768 / 12
Training-time interval	$[0.01, 0.99]$	$[0.01, 0.99]$
Optimizer / learning rate	AdamW / 10^{-4}	AdamW / 10^{-4}
Weight decay / learning-rate scheduler	0 / none	0 / none
Batch size / epochs	128 / 2,000	128 / 2,000
Gradient-norm clipping	10	10
EMA decay / update interval	0.9999 / every step	0.9999 / every step

B.3 FAMILYSPACE CONTROLLABLE SAMPLING

A requested family template $\mathbf{f} = (f_1, \ldots, f_N)$ specifies the family of each source token, not its exact element identity. We independently sample $\mathbf{z}_i^{(0)} \sim r_{f_i}$ and transport the sequence using the EMA flow. The reported implementation uses an Euler grid from $t_0 = 0.01$ to $t_L = 1$ with 100 grid points, giving 99 Euler updates. At each update,

$$\mathbf{Z}_{\ell+1} = \mathbf{Z}_\ell + (t_{\ell+1} - t_\ell)\mathbf{v}_{\theta_{\mathrm{EMA}}}(\mathbf{Z}_\ell, t_\ell), \qquad \ell = 0, \ldots, L-1, \tag{22}$$

where $\mathbf{v}_\theta$ denotes the backbone's sampling-time velocity. The initial state $\mathbf{Z}_0$ is the source sample. After the final Euler update, the network makes one additional clean-endpoint prediction at $t = 1$. The frozen VAE decoder maps this prediction to atomic species, fractional coordinates, and lattice parameters.

C EVALUATION DETAILS

C.1 COMPOSITIONAL AND STRUCTURAL VALIDITY

Following prior work (Miller et al., 2024; Zeni et al., 2025), we evaluate validity from both compositional and structural perspectives. A crystal is considered valid only if it satisfies both criteria.

Compositional validity is evaluated with SMACT using the reduced stoichiometry. A composition is considered valid if at least one oxidation-state assignment satisfies charge neutrality and the Pauling electronegativity rules. Single-element structures are accepted by default, while all-metal alloys are allowed without enforcing ionic charge neutrality.

Structural validity is assessed on the MatterSim-relaxed structures rather than on the structures before relaxation. A structure is considered structurally valid if its minimum pairwise atomic distance under periodic boundary conditions is at least 0.5 Å and its unit-cell volume is at least 0.1 Å^3.

C.2 Uniqueness and novelty

Uniqueness and novelty are evaluated on relaxed structures using pymatgen's `StructureMatcher` (Ong et al., 2013; Zeni et al., 2025). The length, site, and angle tolerances are `ltol=0.2`, `stol=0.3`, and `angle_tol=5` degrees. For MP-20, novelty is evaluated against the Alex–MP reference distributed with MatterGen as `reference_MP2020correction.gz`. For MPTS-52, novelty is evaluated against the 27,380 structures parsed from its training CSV. Uniqueness is computed within the full relaxed samples.

C.3 Relaxation and thermodynamic stability

Generated crystals are relaxed with `MatterSim-v1.0.0-1M` (Yang et al., 2024). Both atomic positions and cell parameters are optimized with FIRE and an `ExpCellFilter`. The stopping threshold is 0.05 eV/Å, with a maximum of 3,000 steps. Finite final energies are processed with `MaterialsProject2020Compatibility` and compared with the MP2020 hull reference for both datasets.

Using the reported energy above the reference hull, we define

$$\mathrm{S} = \{j : E_{\mathrm{hull},j} \le 0\ \mathrm{eV/atom}\}, \qquad \mathrm{MS} = \{j : E_{\mathrm{hull},j} \le 0.1\ \mathrm{eV/atom}\}. \tag{23}$$

S denotes predicted stable structures. MS also includes structures within 0.1 eV/atom of the hull. The combined success sets are

$$\mathrm{SUN} = \mathrm{S} \cap \mathrm{U} \cap \mathrm{N}, \qquad \mathrm{MSUN} = \mathrm{MS} \cap \mathrm{U} \cap \mathrm{N}. \tag{24}$$

C.4 Family-level control metrics

Evaluation setup. For each generated crystal, we compare the requested family composition with the composition decoded by the frozen VAE. Metrics are computed over all generated samples before validity filtering. For sample j, let

$$\mathbf{c}_j = (c_{j,1}, \ldots, c_{j,K}), \qquad \widehat{\mathbf{c}}_j = (\widehat{c}_{j,1}, \ldots, \widehat{c}_{j,K})$$

denote the requested and decoded family-count vectors, respectively, where $c_{j,k}$ and $\widehat{c}_{j,k}$ are the requested and decoded numbers of atoms from family k in sample j. We further define the corresponding family sets as

$$\mathcal{F}_j = \{k : c_{j,k} > 0\}, \qquad \widehat{\mathcal{F}}_j = \{k : \widehat{c}_{j,k} > 0\}.$$

All metrics depend only on composition and are independent of atom ordering.

Family composition accuracy. Let M denote the number of generated samples used to compute the control metrics. *Family-set accuracy* measures the fraction of samples that contain exactly the requested families:

$$R_{\mathrm{set}} = \frac{1}{M} \sum_{j=1}^{M} \mathbb{I}[\mathcal{F}_j = \widehat{\mathcal{F}}_j]. \tag{25}$$

Family-count accuracy additionally requires the number of atoms from every family to match:

$$R_{\mathrm{count}} = \frac{1}{M} \sum_{j=1}^{M} \mathbb{I}[\mathbf{c}_j = \widehat{\mathbf{c}}_j]. \tag{26}$$

For experiments that specify family ratios, we report *Family-ratio accuracy*:

$$R_{\text{ratio}} = \frac{1}{M}\sum_{j=1}^{M}\mathbb{I}\Big[\mathcal{F}_j = \widehat{\mathcal{F}}_j \ \wedge\ \text{red}(\mathbf{c}_j) = \text{red}(\widehat{\mathbf{c}}_j)\Big], \tag{27}$$

where $\text{red}(\cdot)$ divides the nonzero family counts by their greatest common divisor. All three accuracies are reported as percentages.

Target-family fraction. For single-family control, we additionally report the *Target-family fraction*. Let

$$\widehat{p}_{j,k} = \frac{\widehat{c}_{j,k}}{\sum_{\ell=1}^{K}\widehat{c}_{j,\ell}}$$

denote the decoded fraction of family k in sample j. For a target family k, the Target-family fraction is defined as

$$\overline{p}^{(k)}_{\text{target}} = \frac{1}{M_k}\sum_{j=1}^{M_k}\widehat{p}_{j,k}, \tag{28}$$

where M_k is the number of generated samples for target family k. It measures the average fraction of decoded atoms belonging to the requested family and is reported as a percentage.

D Additional Experiments

D.1 Within-family compositional diversity

Family templates specify elemental families but do not fix the exact elements within each family. Therefore, a given multi-family template should allow multiple elemental compositions rather than collapsing to a single element choice for each family. For example, samples generated from the Alkali family include Li, Na, K, Rb, Cs, and H, rather than consistently selecting only Li. To verify that sampling from a specified family can produce diverse elemental choices, we analyze the within-family compositional diversity of the multi-family ratio control experiments in Table 4. As shown in Table 9, each family template produces many distinct elemental compositions. These results show that FamilySpace controls family-level composition without fixing the exact elemental composition within each family combination.

Table 9: Within-family compositional diversity in the multi-family ratio control experiments in Table 4, with 10,000 generated samples for each family combination. *Reduced compositions* column reports the number of distinct reduced elemental formulas, while *Element sets* column reports the number of distinct element combinations, irrespective of stoichiometric ratios.

Family combination	Reduced compositions	Element sets
TM–Ch	4,337	1,950
Alkali–TM–Ch	2,418	1,284
AE–TM–Ch	3,004	1,885
RE–TM–Ch	2,435	1,899
Alkali–TM–Pn–Ch	1,013	692

D.2 Stability-oriented exploration protocols

In Section 5.2, we use selected family templates for stability-oriented exploration. We first choose six chemically motivated templates whose family compositions are commonly found in reported materials (Sirdeshmukh et al., 2001; Kumta & Risbud, 1994; Zinkevich, 2007; Mitchell & Ibers, 2002; Hargittai, 2000). As an initial exploration, we generate 10,000 structures using a mixture of these six templates. The samples are distributed approximately uniformly across the six family ratios: four templates receive 1,667 samples each and two receive 1,666 samples each.

The results are summarized in Table 10. Among the six templates, Alkali–Halogen 1:1 and RE–Halogen 1:3 achieve the two highest MSUN rates. We therefore select these two templates for

further evaluation and independently generate 10,000 additional structures for each. The focused runs confirm that both templates produce a high proportion of metastable candidates.

Table 10: Stability-oriented exploration on MP-20. N denotes the number of generated samples.

Family template	N	UN (%)	S (%)	MS (%)	SUN (%)	MSUN (%)
Training templates (reference)	10,000	69.53	6.26	42.99	1.38	15.31
Six-template mixture						
Alkali–Halogen 1:1	1,667	51.71	14.70	90.16	1.74	42.11
AE–Halogen 1:2	1,667	62.75	9.90	70.13	0.42	34.43
Alkali–Ch 2:1	1,667	46.37	21.96	86.62	2.10	34.43
AE–Ch 1:1	1,667	19.80	22.50	77.38	0.12	4.56
RE–Ch 2:3	1,666	78.26	9.01	36.65	0.93	16.17
RE–Halogen 1:3	1,666	77.99	8.17	70.77	6.10	54.43
Independent focused runs						
Alkali–Halogen 1:1	10,000	45.25	13.56	90.23	0.70	36.12
RE–Halogen 1:3	10,000	63.00	7.87	70.36	4.97	41.50

D.3 Application-oriented exploration

For application-oriented exploration, we generate 10,000 structures for each of the five family templates in Table 11. All generated structures are evaluated using the same relaxation, hull-energy calculation, and reference matching procedures described in Appendix C. We retain candidates satisfying $\mathrm{MSUN} = \mathrm{MS} \cap \mathrm{U} \cap \mathrm{N}$ and examine whether their chemical compositions have been reported in the literature for relevant applications.

Table 11: Application-oriented exploration on MP-20. We generate 10,000 structures for each family template. U, N, UN, S, SUN, MS, and MSUN are reported as percentages. TM denotes Transition metal, AE Alkaline earth, RE Rare earth, Ch Chalcogens, and Pn Pnictogens, following Appendix A.1.

Family template	U (%)	N (%)	UN (%)	S (%)	SUN (%)	MS (%)	MSUN (%)
TM–Ch	90.25	89.01	85.93	2.34	1.07	20.86	9.79
Alkali–TM–Ch	87.63	88.93	84.30	1.12	0.18	21.17	10.35
AE–TM–Ch	82.64	85.27	78.98	1.54	0.08	23.94	7.31
RE–TM–Ch	87.37	89.48	83.84	1.34	0.36	21.64	9.97
Alkali–TM–Pn–Ch	94.85	98.73	94.40	0.05	0.04	13.74	10.54

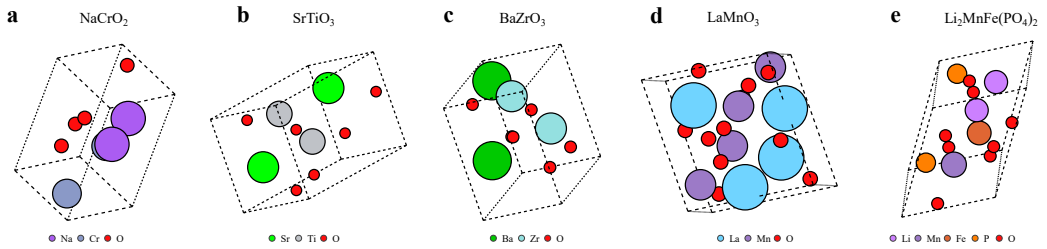


Figure 5: Representative generated MSUN candidates with application-relevant compositions. (a) $\mathrm{NaCrO_2}$ from the Alkali–TM–Ch template; (b) $\mathrm{SrTiO_3}$ and (c) $\mathrm{BaZrO_3}$ from the AE–TM–Ch template; (d) $\mathrm{LaMnO_3}$ from the RE–TM–Ch template; and (e) $\mathrm{Li_2MnFe(PO_4)_2}$ from the Alkali–TM–Pn–Ch template.

Figure 5 shows five representative MSUN candidates whose compositions are associated with reported functional materials. The $\mathrm{NaCrO_2}$ composition in panel (a) has been investigated as a sodium-ion battery cathode, with nitrogen-doped carbon coating used to improve its rate capability and cycling stability (Cai et al., 2025). The compositions in panels (b) and (c), $\mathrm{SrTiO_3}$ and

$BaZrO_3$, are associated with photocatalytic overall water splitting (Luo et al., 2025a) and proton-conducting ceramic electrolyte hosts (Tang et al., 2025), respectively. For $LaMnO_3$ in panel (d), recent work has investigated unconventional cubic nanosheets for oxygen-evolution electrocatalysis (Dai et al., 2025). The $Li_2MnFe(PO_4)_2$ composition in panel (e) is stoichiometrically equivalent to $LiMn_{0.5}Fe_{0.5}PO_4$, which has been studied as a lithium-ion battery cathode with engineered carbon coatings (Hou et al., 2025).