# Fast Organic Crystal Structure Prediction with Unit Cell Flow Matching

**Alston Lo**[1,*] **Luka Mucko**[2,3,4,*] **Austin H. Cheng**[3,4,*] **Andy Cai**[3]
**Alastair J. A. Price**[3] **Wojciech Matusik**[1] **Alán Aspuru-Guzik**[3,4,5,6,7]

[1]MIT [2]FER, University of Zagreb [3]University of Toronto [4]Vector Institute
[5]Acceleration Consortium [6]CIFAR [7]NVIDIA

## Abstract

Organic crystal structure prediction (CSP) is a requirement for computational modelling of organic solids, but traditionally costs several CPU-years per molecule. Generative models such as OXtal dramatically reduce this cost by sampling stable organic crystal structures directly. However, OXtal forgoes explicit lattice parametrization in favour of modelling large crops of the bulk material with expensive triangle layers, which can incur a computational cost of minutes per molecule. In this paper, we reduce this to seconds with CLARI, a large-scale flow matching model that generates redundancy-free unit cells and replaces triangle layers with pair-bias attention. CLARI requires only atom types and bonds as input and does not need an RDKit-sanitizable input molecule, which expands its applicability to challenging chemistries such as fullerenes, metal complexes, and atom clusters. We further ablate key design choices such as auxiliary losses, timestep distributions, noise priors, and self-conditioning. On OXtal's test sets, we surpass OXtal's solve rate while obtaining a speedup of 15–30×. Because CLARI also models explicit hydrogens, it supports inference-time scaling via direct energy ranking, without any decoration or relaxation step. When generating 150 crystals and selecting the top-30 by energy, we further improve solve rate while maintaining a speedup of 5–8×. We also introduce the CSD Teaching Subset as a new test split of diverse and complex molecules for future benchmarking. Our contributions enable CSP within seconds, making large-scale virtual screening of organic solids practical.

Code: https://github.com/the-matter-lab/clari

## 1 Introduction

The properties of organic solids depend strongly on crystal packing. Knowing a molecule's crystal structure is therefore a prerequisite for modelling solid-state properties in applications spanning fertilizers [1], pesticides [2], pigments [3, 4], food [5], energetic materials [6], pharmaceuticals [7, 8], and organic electronics [9] such as organic light-emitting diodes [10] and organic photovoltaics, as well as emerging flexible materials [11, 12]. Crystal structures also enable the design of templates that seed desired crystal forms of other molecules [13, 14]. Yet in many practical settings, no experimental crystal structure is available, so it must be predicted computationally.

The problem of crystal structure prediction (CSP) has challenged scientists for decades [15], owing to the large and rugged search space of molecular crystals. Traditional CSP pipelines tackle this by executing a dense computational funnel. Starting from a large set of randomly generated candidate structures, they iteratively filter and re-rank candidates using successively more accurate but costly quantum chemistry methods, culminating in density functional theory (DFT) relaxations [16–18]. While these methods are accurate [19], exhaustive search pipelines can require several CPU-years

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

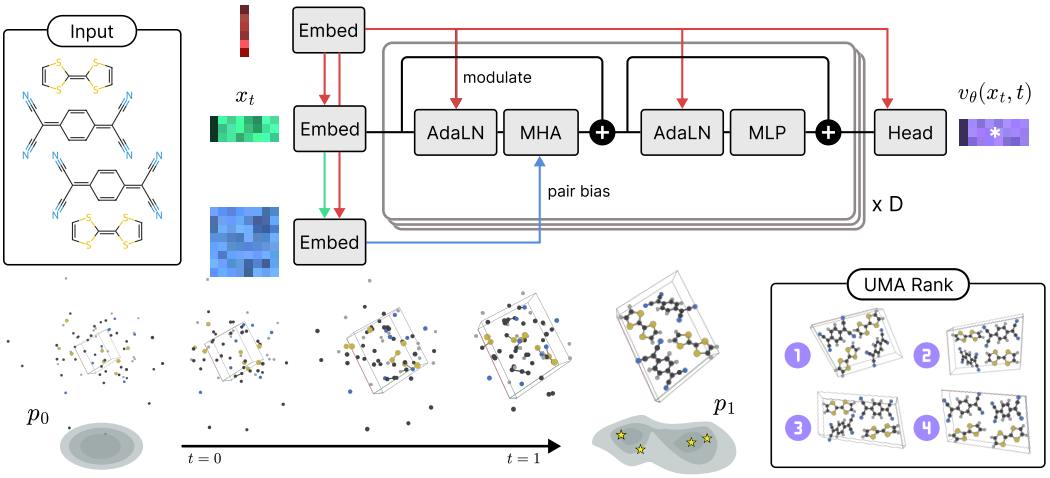



Figure 1: Overview of CLARI. Given a molecular graph as input, CLARI directly generates a single unit cell comprising atom coordinates and lattice vectors via a flow matching trajectory. No bulk expansion or triangle layers are required.

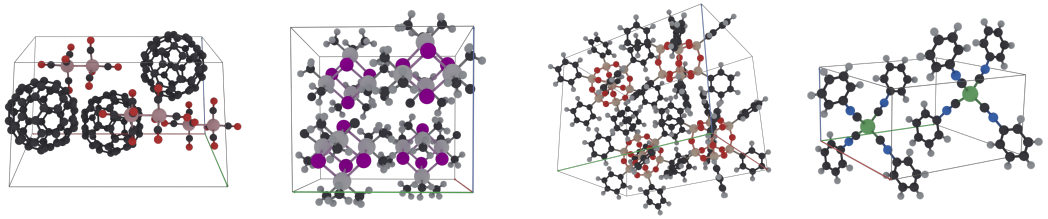
per molecule [20, 21], making them prohibitive for screening even modestly sized virtual libraries. In practice, this bottleneck confines traditional CSP to targeted pharmaceutical campaigns [22] and renders large-scale screening on the basis of solid-state properties all but infeasible [23].

A given molecule can form multiple stable crystal forms in a phenomenon known as *polymorphism* [24]. The one-to-many relationship between chemical identity and crystal polymorphism makes the problem map closely to generative modelling. Indeed, recent work has introduced generative models as a direct alternative to energy-based search. OXtal [25] demonstrated that a generative model can learn to sample the distribution of experimentally observed organic crystal structures end-to-end. While paradigm-shifting, OXtal inherits a number of expensive design choices from AlphaFold3 [26]: it represents crystals in bulk form, processing multiple symmetry-related copies of each molecule simultaneously, and it relies on triangle-update layers [27] to propagate pairwise geometric information. Together, these choices impose a large computational footprint that makes training and inference costly, and poses barriers to screening ultra-large libraries.

In this work, we show that two targeted refinements suffice to narrow the gap between generative CSP and practical screening. First, rather than modelling a bulk crop of symmetry-equivalent copies, we train CLARI directly on *unit cells*, jointly predicting atom coordinates and lattice vectors under a flow matching objective (Figure 1). Second, we *replace triangle-update layers* with a pair-bias attention architecture, which transmits pairwise geometric information through attention logits without the cubic cost of triangle operations. Forgoing bulk crops and triangle layers may make global consistency harder to learn, but the resulting gain in efficiency enables more training updates and faster inference. Together, these choices allow CLARI to surpass OXtal in prediction quality while also accelerating sampling by roughly 15–30$\times$ on the CSP Blind Tests, or 5–8$\times$ end-to-end when ranking by energy with the Universal Model for Atoms (UMA) [28]. We also train on all-atom structures including hydrogens, which enables direct energy-based ranking without any decoration or relaxation step. Ablations isolate the contribution of each component: self-conditioning, the choice of lattice source distribution, auxiliary losses, and the timestep schedule.

In summary, our main contributions are:

- We introduce CLARI, a flow matching model for organic CSP that operates on a single unit cell with a pair-bias attention architecture, sampling roughly 15–30$\times$ faster than OXtal on the CSP Blind Tests (5–8$\times$ end-to-end with UMA energy ranking). Inference takes on the order of seconds per molecule. Unlike OXtal, CLARI does not require input conformers, which can be cumbersome to obtain for complex chemistries like metal complexes and atom clusters (Figure 2).
- The resulting speed opens generative CSP to virtual library screening at scale [29], a regime previously accessible only to fast but coarse traditional heuristics. This throughput also facilitates



(a) YUGWUT (b) VISMOB (c) DORRAF (d) XADQUP

Figure 2: Crystal structures predicted by CLARI across chemically diverse classes: (a) YUGWUT, a $C_{60} \cdot Co_2(CO)_8$ fullerene cocrystal; (b) VISMOB, the $[(CH_3)_3PtI]_4$ trimethylplatinum iodide tetramer with a $Pt_4(\mu_3\text{-}I)_4$ cubane core; (c) DORRAF, an octaphenyl-substituted $Si_8O_{12}$ POSS cage; (d) XADQUP, a tetrahedral transition-metal complex.

inference-time scaling, including energy-based ranking via joint heavy-atom and hydrogen modelling.

- For a rigorous final evaluation, we combine complementary metrics into a single evaluation suite and introduce a new test split drawn from the CSD Teaching Subset, covering chemically complex systems including fullerenes, boranes, and organometallic complexes that were excluded from or underrepresented in prior benchmarks. We also conduct systematic ablations of architectural and training choices, revealing which components drive performance gains.

## 2 Related work

**Computational (simulation-based) organic crystal structure prediction.** Organic CSP has historically been framed as a two-stage pipeline consisting of structure generation followed by energy ranking. This paradigm has been formalized and benchmarked through the Cambridge Crystallographic Data Centre (CCDC) Blind Tests [16, 17, 21, 30–34], which evaluate CSP methods on unpublished experimental structures. Early approaches relied on extensive exploration of the configurational space using quasirandom sampling [35, 36], simulated annealing [37], and evolutionary algorithms [38]. While these methods are general, they require generating and evaluating a large number of candidate structures, making CSP computationally expensive in practice.

**Machine learning for accelerating CSP.** Subsequent work has focused on reducing the cost of the ranking stage rather than altering the search paradigm. In particular, machine learning interatomic potentials (MLIPs) have been widely adopted as surrogates for DFT, significantly accelerating energy and force evaluations [17]. Systems such as FastCSP [39] demonstrate that MLIPs can speed up traditional pipelines by orders of magnitude. However, these approaches still filter large candidate sets, leaving the combinatorial burden of structure generation largely unchanged.

**Generative CSP.** More recently, generative approaches have emerged as an alternative to exhaustive search, aiming to directly sample low-energy crystal structures. OXtal [25] introduced this paradigm for organic CSP using an AlphaFold3-inspired architecture. Follow-up work explores different generative formulations, including reinforcement learning in PackFlow [40] and flow matching on rigid bodies in MolCrystalFlow [41]. These methods demonstrate that sufficiently expressive generative models can substantially reduce the reliance on downstream ranking. However, current approaches either employ computationally intensive architectures (e.g., triangle updates over large crops) or rely on restrictive representations such as rigid molecules, limiting efficiency and applicability.

**Generative models for chemistry.** Outside of organic CSP, parallel work in inorganic materials has developed a range of generative models for periodic crystals, beginning with variational approaches such as CDVAE [42] and followed by models that more explicitly incorporate symmetry and periodicity [43–49]. A consistent design choice across these methods is the use of *explicit lattice parametrizations*, in which the unit cell and atomic positions are modelled directly. Our work is also related to advances in generative modelling of molecular geometry, including proteins [50–56], biomolecular complexes [26, 57], and molecules [58–63]. These works demonstrate the effectiveness of generative models for structured 3D data, from which we borrow several techniques. LLM-based

approaches such as CrystaLLM [64] instead generate CIF files directly. This representation is flexible and expressive (e.g., for space groups and disorder), but can be considerably less compact than unit cell point clouds, which may limit scalability and efficiency.

# 3 CLARI

## 3.1 Crystal structure prediction

We represent a crystal unit cell as a tuple $(\boldsymbol{L}, \boldsymbol{C}, \boldsymbol{F}, \boldsymbol{E})$, where $\boldsymbol{L} \in \mathbb{R}^{3\times3}$ is a lattice matrix with primitive vectors stored as rows, $\boldsymbol{C} \in \mathbb{R}^{N\times3}$ are Cartesian atom coordinates with zeroed centroid, $\boldsymbol{F} \in \mathbb{R}^{N\times d}$ are atom features such as atomic numbers, and $\boldsymbol{E} \in \mathbb{R}^{N\times N}$ is an adjacency matrix of bond types. Unit cells periodically tile the space to produce full crystals through translations by $\boldsymbol{L}^\top \boldsymbol{z} \in \mathbb{R}^3$, for any $\boldsymbol{z} \in \mathbb{Z}^3$. This defines a notion of *periodic* distance $d_{\boldsymbol{L}}(\boldsymbol{p}, \boldsymbol{q}) = \min_{\boldsymbol{z}\in\mathbb{Z}^3} ||\boldsymbol{p} - \boldsymbol{q} - \boldsymbol{L}^\top \boldsymbol{z}||_2$, which is the minimum distance between the periodic images of points $\boldsymbol{p}$ and $\boldsymbol{q}$ in the unit cell.

The problem of crystal structure prediction is to infer the positions $(\boldsymbol{L}, \boldsymbol{C})$ given the crystal molecular graph $(\boldsymbol{F}, \boldsymbol{E})$. Concretely, the input is the 2D molecular graph of the molecules in the unit cell with multiplicity; we do not consider the space group as input. We frame CSP as a conditional generative modelling problem and adopt a flow matching approach due to its simplicity and recent success in biomolecular modelling [65]. A key design choice is that we generate $(\boldsymbol{L}, \boldsymbol{C})$ as a unified object by treating the primitive lattice vectors as three additional *virtual* points. That is, we concatenate them row-wise into a matrix $\boldsymbol{x} = (\frac{1}{2}\boldsymbol{L}; \boldsymbol{C})/\sigma \in \mathbb{R}^{(3+N)\times3}$ with $\sigma$ chosen to normalize $\boldsymbol{x}$ to roughly unit variance across the dataset. Assuming $(\boldsymbol{F}, \boldsymbol{E})$ is fixed, $\boldsymbol{x}$ is invariant under (i) signed permutations of the lattice rows, (ii) permutations of atomic rows that are also automorphisms of the crystal molecular graph, (iii) rotations, and (iv) independent periodic translations of any connected molecular subgraph, which we call a *body* or *component*.

## 3.2 Flow matching

Flow matching [66–68] learns a continuous-time vector field that transports a tractable source distribution $p_0$ into the data distribution $p_1$. Adopting the linear interpolant $\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$, a network $v_\theta$ is trained to regress the target velocity:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t,\,\boldsymbol{x}_0,\,\boldsymbol{x}_1}\left[\left\|v_\theta(\boldsymbol{x}_t, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\right\|^2\right], \quad t \sim p(t),\ \boldsymbol{x}_0 \sim p_0(\boldsymbol{x}),\ \boldsymbol{x}_1 \sim p_1(\boldsymbol{x}). \tag{1}$$

In practice, we split $\mathcal{L}_{\text{FM}}$ into lattice and coordinate terms that are summed $\mathcal{L}^{\boldsymbol{L}}_{\text{FM}} + \mathcal{L}^{\boldsymbol{C}}_{\text{FM}}$. Each term is computed using a mean-squared error so that they contribute equal weight to the final loss. A trained flow matching model $v_\theta$ can then be used to generate samples by evolving $\boldsymbol{x}_0 \sim p_0(\boldsymbol{x})$ from the source distribution according to $\mathrm{d}\boldsymbol{x}_t = v_\theta(\boldsymbol{x}_t, t)\mathrm{d}t$ using any ordinary differential equation (ODE) solver.

**Time distribution and schedule.** A rich and important design space lies in choosing the timestep training distribution $p(t)$ and sampling discretization $(t_i)_i$. Prior work has found that skewing towards later timesteps leads to better results in biomolecular modelling [55], where fine-grained local details matter more than, for example, in images. However, we argue that crystals also require attention to the overall global arrangement. In Section 4.3, we explore a variety of options including uniform, logit-normal [69], and beta [55, 63] distributions and linear and log discretizations (Figure 3a).

**Source distribution.** The de facto choice of $p_0$ is a unit normal distribution. However, a random normal matrix with high probability yields unrealistic cell densities and degenerate periodic images, which we find leads to a considerable degradation in performance. Instead, we use a data-informed prior $p_0$ by decomposing the lattice into three components (atom density, cell angles, and volume-normalized cell lengths) and fitting a Gaussian distribution to each using the training data (Figure 3b). To sample from $p_0$, we sample each component independently, reconstruct the lattice matrix, and then apply a random rotation and signed permutation. Further details are given in Appendix B.1.

**Auxiliary losses.** In order to improve physical quality, we enrich the base flow matching objective with auxiliary losses between the one-step estimate $\hat{\boldsymbol{x}}_1 = \boldsymbol{x}_t + (1-t)\, v_\theta(\boldsymbol{x}_t, t)$ and the ground truth $\boldsymbol{x}_1$. In protein models, it is common to use a bond loss, smooth local distance difference test (LDDT), or distogram loss [26, 55]. We propose a new pair of losses relevant for crystal generation: (i) $\mathcal{L}_{\text{vol}}$, the relative error between predicted and target lattice volumes; and (ii) $\mathcal{L}_{\text{pair}}$, a pairwise periodic distance error that also penalizes steric clashes. The full objective is then $\mathcal{L}_{\text{CLARI}} = \mathcal{L}^{\boldsymbol{L}}_{\text{FM}} + \mathcal{L}^{\boldsymbol{C}}_{\text{FM}} + \mathcal{L}_{\text{vol}} + \mathcal{L}_{\text{pair}}$. The formal definitions of the auxiliary losses are provided in Appendix B.4.

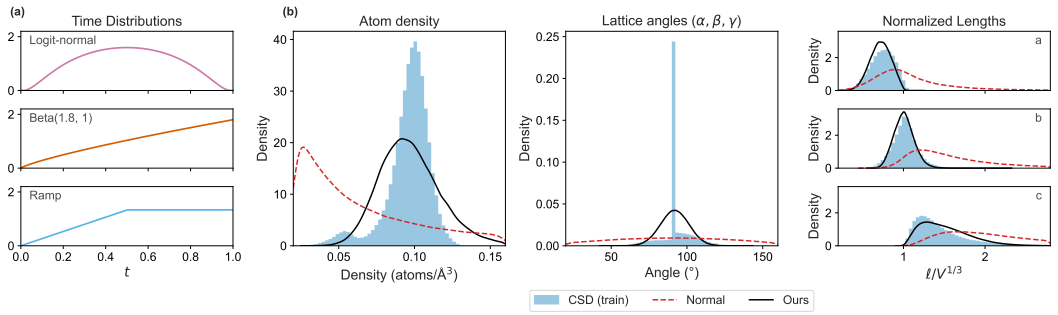



Figure 3: **(a)** Time distributions $p(t)$ considered in our ablations: uniform (not depicted), logit-normal, Beta$(1.8, 1)$, and ramp. **(b)** Marginals of the lattice source distribution $p_0$ versus the Cambridge Structural Database (CSD) training set: atom density (atoms/Å$^3$), the three unit cell angles, and the sorted volume-normalized lattice lengths $\ell/V^{1/3}$. A standard Gaussian source concentrates mass at vanishing density, producing degenerate lattices, and badly misfits the length distribution. Our fitted prior closely tracks the data marginals.

**Self-conditioning.** CLARI also uses self-conditioning [58, 60, 70] whereby the model recycles its own endpoint estimate $\hat{\boldsymbol{x}}_1$ from the previous timestep as input. This incurs effectively no cost during inference, but slows training by $\sim$20% due to an extra forward pass (on a half batch) every step. Despite this, we find self-conditioning to yield a consistent performance improvement, so we include it in our final models.

### 3.3 Architecture

An emerging paradigm in biomolecular structure modelling is using large scalable Transformers that are not inherently equivariant but instead learn it through large-scale training with data augmentation [26, 55, 59]. In the same spirit, CLARI implements $v_\theta$ using the Diffusion Transformer (DiT) [71] architecture with modern Transformer [72] tricks such as gated attention [73], QKNorm [74], and SwiGLU [75]. Since the lattice is treated as virtual points, the lattice and atom tokens participate identically throughout the network and no special pooling layers are required. Pair features are created from bond, topological and geometric distance information, which modulate the DiT blocks through additive pair attention biasing. Finally, AdaLN-Zero blocks [71] are used to modulate the DiT with global-level information such as the timestep, lattice, and molecular formulae. Importantly, CLARI contains no triangular operations, which results in a significantly more scalable and efficient architecture. This allows CLARI to scale to 350M parameters even for unit cells with up to 512 atoms. Further details are in Appendix B.2.

### 3.4 Optimal transport and augmentation

Standard flow matching draws $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ independently, but coupling them via an approximate optimal-transport map straightens trajectories and reduces variance of the regression target [76–78]. We aim to align $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ with respect to (i) signed lattice permutation, (ii) atomic permutation, and (iii) rotations, as described in Section 3.1. However, an exact coupling under the joint action is intractable. We therefore approximately align $\boldsymbol{x}_1$ to $\boldsymbol{x}_0$ by composing three tractable steps that each minimize squared deviation along individual symmetry groups: brute-force signed-permutation lattice alignment, atom-permutation alignment via the Hungarian algorithm [79], and a final weighted Kabsch [80] pose alignment. We refer readers to Appendix B.5 for further details. Note that alignment effectively performs augmentation due to the invariance of our prior with respect to these actions. We handle the last remaining symmetry of periodic translation (iv) through an augmentation step before alignment. Specifically, we randomly translate the unit cell, re-wrap the centroid of each component to lie in $[0, 1)^3$, and finally re-centre the centroid of the atomic coordinates.

### 3.5 Inference-time scaling

For an efficient model like CLARI, a simple way to improve performance is inference-time scaling, also known as test-time compute [81]. Recent work [57] has shown the value of these techniques for flow-based atomistic generation, applying methods like beam search and Feynman-Kac steering [82, 83] to binder design. We leave a principled exploration of such methods for future work and, for

now, opt for a simple best-of-$N$ approach. Specifically, we generate multiple candidate structures per target, score them with the UMA model `uma-s-1p2` [28], and report results on a subset of lowest-energy structures. Because CLARI generates full all-atom crystals, we can immediately rank candidates by energy without having to perform any hydrogen decoration or relaxation steps.

# 4 Experiments

We evaluate CLARI on held-out crystals (*Rigid* and *Flexible* OXtal subsets), three crystal structure prediction (CSP) Blind Tests (CSP5–7), and the Cambridge Structural Database (CSD) Teaching Subset. We begin by comparing CLARI against baselines in Section 4.2. Then, we motivate our design choices through a series of ablations that build up to CLARI in Section 4.3.

## 4.1 Experimental setup

**Dataset.** We train CLARI on the Cambridge Structural Database (CSD) [84], a million-scale database of experimentally determined organic and metal-organic crystal structures. We keep entries that have a 3D structure, are non-polymeric, and are derived from single-crystal diffraction at ambient pressure. For non-test splits, we further filter for entries deposited up to May 1, 2025 (following Jin et al. [25]), with R-factor below $10\%$, and at most $512$ atoms in the unit cell. Unlike prior efforts [25, 40], we do not remove hydrogens or require RDKit sanitization, allowing us to retain a much larger set of entries. After metadata filtering, we found the dataset to still be relatively noisy due to unflagged polymers or superimposed molecules from improperly specified disorder. We attempt to catch these cases using distance-based thresholds, which are described in Appendix A among other details. The OXtal test crystal families and the CSD Teaching Subset are combined to create our held-out test pool.

**Training and inference.** We train a medium model CLARI-M (88M parameters) for 150,000 steps on $4$ NVIDIA H100 GPUs. We then scale the model to CLARI-L (173M parameters) on $8$ H100 GPUs and further to CLARI-H (350M parameters) on $16$ H100 GPUs for final comparisons against baselines. Hyperparameters are given in Appendix B.6. During training, refcode families are sampled uniformly to avoid bias towards molecules with many polymorphs. We take the representative with the lowest R-factor from each refcode family for validation and testing. For inference, we use the Heun sampler [85] with 20 and 50 steps in Sections 4.2 and 4.3, respectively.

**Metrics.** Robustly assessing predicted crystals is difficult [86], so we report a battery of complementary metrics. For baseline comparisons, we follow the OXtal evaluation protocol with COMPACK [87], which searches for matching clusters of molecules between the predicted and ground-truth crystals. Our primary metric is approximate solve rate **Sol@*k*** ($\uparrow$): a target counts as solved if at least one of its $k$ samples matches the ground truth in at least $8$ of $15$ molecules with $\mathrm{RMSD}_{15} < 2\,\text{Å}$ and no inter-body clashes, where $\mathrm{RMSD}_{15}$ is the root-mean-square deviation across the matched 15-molecule cluster. Throughout the paper, Sol refers to this $\geq 8/15$ variant, matching OXtal for direct comparison. The strict $15/15$ variant is reported in Table 6. For ablations, we use cheaper metrics: (1) **Clash Rate** ($\%$, $\downarrow$), the fraction of structures with an inter-body atom pair closer than their sum of covalent radii; (2) **PoseBusters** pass rate ($\%$, $\uparrow$) [88] on applicable fragments; (3) relative **Volume Error** ($\%$, $\downarrow$); and (4) **EMD PDD** ($\downarrow$), an earth mover's distance (EMD) between pointwise distance distributions (PDD) [89], an isometry invariant of periodic point sets. We split these into *quality* (Clash Rate, PoseBusters) and *reconstruction* (Vol. Error, EMD PDD) metrics.

## 4.2 Comparison to baselines

Leveraging the fast generation speed of CLARI, we apply inference-time scaling through best-of-$N$ sampling (Section 3.5). For each target crystal, we generate $n_s > k$ candidates, rank them by UMA energy without geometry optimization, and retain the $k$ lowest-energy candidates (equivalently, the top-$k$ candidates) to check Sol@$k$. Figure 4 shows that Sol@$k$ increases sublinearly with both the sampling budget $n_s$ and the retained set size $k$. However, for fixed $k$, increasing $n_s$ does not always improve Sol@$k$, indicating ranking noise. Thus, we tune $n_s$ with respect to Sol@30 on a held-out set of $50$ crystals drawn from the CSD Teaching Subset, minimizing leakage onto the OXtal benchmark (Figure 8). We find that $n_s = 150$ works best.

Table 1 reports Sol across all subsets. To estimate uncertainty, we generate a pool of $1000$ candidates per target and compute bootstrap estimates of Sol@$k$ using $5000$ resamples. In each resample, we draw $n_s$ candidates with replacement from the $1000$-candidate pool, proceed with the above ranking procedure to compute Sol@$k$, and then average Sol@$k$ over resamples. We report bootstrap standard

Table 1: Approximate solve rate Sol@$k$ across the OXtal benchmark subsets. $n_s$ is the generation budget (samples drawn from the model), while $k$ is the number used to compute Sol@$k$. $\text{DFT}_{\text{avg}}$ is the average across participants in the corresponding CSP Blind Test as evaluated by Jin et al. [25] (they use $n_s = k = 464/83/868$ for CSP5/CSP6/CSP7, respectively). When $k < n_s$, our methods select the top-$k$ by UMA energy. Best per column within each block of our rows in **bold**. Bootstrap means over 5000 resamples are reported, and standard errors are given in Table 7.

| Method | $n_s$ | $k$ | Rigid (50) | Flexible (50) | CSP5 (6) | CSP6 (5) | CSP7 (8) | Teach. (773) |
|---|---|---|---|---|---|---|---|---|
| $\text{DFT}_{\text{avg}}$ | – | – | – | – | 0.544 | 0.496 | 0.421 | – |
| OXtal | 30 | 30 | 0.300 | 0.220 | 0.167 | 0.200 | 0.125 | – |
| CLARI-M | 30 | 30 | 0.697 | 0.241 | 0.554 | 0.311 | 0.210 | 0.442 |
| CLARI-L | 30 | 30 | 0.731 | 0.287 | 0.681 | 0.355 | 0.245 | 0.461 |
| CLARI-H | 30 | 30 | 0.766 | 0.378 | 0.705 | 0.445 | 0.280 | 0.584 |
| CLARI-M | 150 | 30 | 0.712 | 0.260 | 0.616 | 0.374 | 0.225 | 0.470 |
| CLARI-L | 150 | 30 | 0.772 | 0.346 | **0.789** | 0.480 | 0.263 | 0.484 |
| CLARI-H | 150 | 30 | **0.798** | **0.479** | 0.785 | **0.612** | **0.393** | **0.621** |
| CLARI-M | 400 | 200 | 0.879 | 0.506 | 0.863 | 0.657 | 0.384 | 0.646 |
| CLARI-L | 400 | 200 | **0.919** | 0.596 | **0.975** | 0.729 | 0.566 | 0.669 |
| CLARI-H | 400 | 200 | 0.891 | **0.678** | 0.965 | **0.766** | **0.822** | **0.771** |
| CLARI-M | 1000 | 1000 | **0.960** | 0.620 | **1.000** | **0.800** | 0.500 | 0.754 |
| CLARI-L | 1000 | 1000 | 0.940 | 0.760 | **1.000** | **0.800** | 0.875 | 0.763 |
| CLARI-H | 1000 | 1000 | **0.960** | **0.780** | **1.000** | **0.800** | **1.000** | **0.856** |

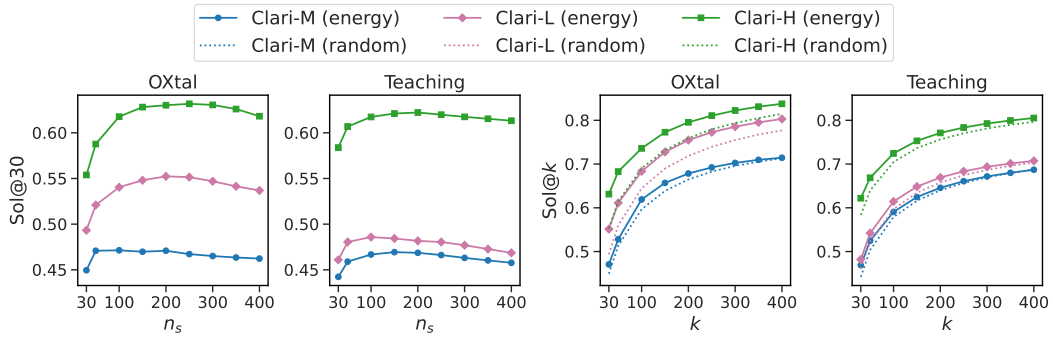



Figure 4: Inference-time scaling on OXtal's aggregated test set and the CSD Teaching Subset. *Left:* Sol@30 as a function of the sampling budget $n_s$, with the top 30 samples per target selected by UMA energy. Scaling $n_s$ does not always improve Sol@30, indicating ranking noise. *Right:* Sol@$k$ as a function of the selection size $k$. Solid lines select top-$k$ by UMA energy with $n_s = \max(200, 2k)$, while dotted lines draw $n_s = k$ samples. Selecting top-$k$ by UMA energy provides an advantage over random selection. Bootstrap means over 5000 resamples are plotted.

errors in Table 7. The $n_s = k = 1000$ setting uses the full candidate pool and is therefore reported without uncertainty.

In the setting of $n_s = k = 30$, which disables UMA ranking and matches OXtal's sample count, CLARI-M outperforms OXtal across all subsets. Increasing the scale of CLARI-M to CLARI-L to CLARI-H provides monotonic improvements. Inference-time scaling further improves performance, and CLARI-H with top-30 energy selection achieves the strongest performance among our methods on every subset except CSP5. When sampling $n_s = 400$ candidates and retaining the top-200 by UMA energy, CLARI-L and CLARI-H surpass the average DFT participant in CSP5 (top-464) and CSP7 (top-868) despite retaining fewer candidates. Averaged across OXtal's test set, CLARI-H takes only 2.8 seconds per molecule to generate 150 crystal structures, or 6.6 seconds when downselecting to 30 with energy ranking (Figure 5).

**CSD Teaching Subset.** To evaluate CLARI on more realistic scenarios, we benchmark on the CSD Teaching Subset [90], a collection of >1000 diverse crystals curated for educational purposes, covering

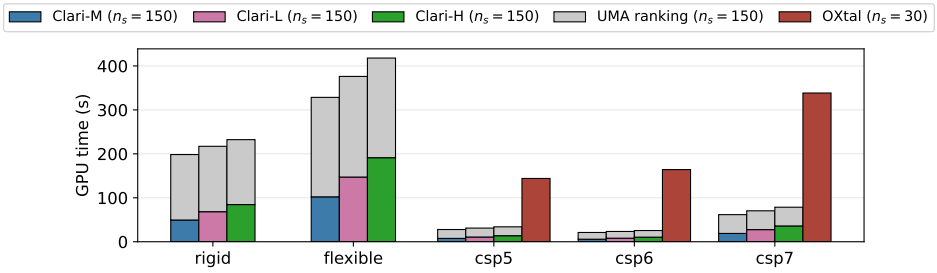



Figure 5: Per-dataset GPU wall-clock time on H100 for $n_s = 150$, where we sample and rank 150 candidates per target. Coloured bars are CLARI sampling; the grey extension is the additional cost of UMA energy ranking on the 150 generated samples. OXtal timings (CSP5–7 only) are approximated for H100 bf16 by dividing the reported L40S timings by a factor of 2.5.

Table 2: Ablation of design choices on the validation set. The first row reports the quality metrics on the Cambridge Structural Database (CSD) validation set for reference. We emphasize that quality metrics are designed not to be blindly optimized but instead to provide another dimension of characterization, as even the ground truth does not obtain a perfect score. PoseBusters is computed on the 71.3% of applicable validation crystals. CLARI-M is obtained from **D** by using Beta$(1.8, 1)$ for $p(t)$ instead of the uniform distribution. Bootstrap means with standard errors are reported over 5000 resamples.

| Model | Clash Rate | PoseBusters | Vol. Error | EMD PDD |
|---|---|---|---|---|
| CSD (Val.) | 0.80 | 92.42 | – | – |
| A (Base DiT) | 25.00 ± 0.50 | 78.14 ± 0.42 | 2.07 ± 0.05 | 11.01 ± 0.05 |
| B (A + lattice tokens) | 23.25 ± 0.49 | 79.04 ± 0.41 | 2.22 ± 0.05 | 11.09 ± 0.05 |
| C (B + auxiliary losses) | 20.33 ± 0.47 | 80.11 ± 0.40 | 1.79 ± 0.04 | 10.37 ± 0.04 |
| D (C + self-cond.) | 9.79 ± 0.35 | 85.23 ± 0.35 | 1.55 ± 0.04 | 9.66 ± 0.03 |
| E (D + normal $p_0$) | 9.32 ± 0.33 | 84.89 ± 0.36 | 2.88 ± 0.06 | 10.71 ± 0.05 |
| CLARI-M | 9.56 ± 0.35 | **87.34 ± 0.34** | 1.59 ± 0.04 | 9.56 ± 0.03 |
| CLARI-L | **7.69 ± 0.32** | 85.89 ± 0.39 | 1.50 ± 0.04 | 9.28 ± 0.03 |
| CLARI-H | 10.81 ± 0.38 | 86.53 ± 0.40 | **1.39 ± 0.03** | **9.13 ± 0.03** |

a wide range of functional groups, valence-shell electron-pair repulsion (VSEPR) structure types, organometallic chemistry, flexible cycles, and stereochemistry. We apply the same standard filters as used for training (R-factor below 10%, at most 512 atoms in the unit cell), yielding 773 crystals. At $n_s = k = 1000$, which represents how CLARI might be used in exhaustive search workflows, CLARI-L attains a Sol@1000 of 0.763 and CLARI-H attains 0.856, demonstrating CLARI's ability to handle a wide space of chemistry. Figure 9 shows that CLARI can make predictions on complex molecules like metal complexes, fullerenes, and atom clusters, which are not RDKit-sanitizable and for which it would be nontrivial to produce the required input conformers for OXtal. Extended metrics are reported in Appendix D.1.

### 4.3 Ablations

To better understand the performance of CLARI, we ablate key design choices on the validation set. We generate 20 samples per crystal and report bootstrap means over 5000 resamples. Each resample draws 5 values per crystal with replacement, aggregates them into a per-crystal score (using a mean for quality metrics and a minimum for reconstruction), and then averages across crystals. Table 2 provides our main results, whereas Table 3 provides ablations across time distribution and discretization. Apart from the scaled-up CLARI-L and CLARI-H, all models in this section use the hyperparameters of CLARI-M, which are listed in Appendix B.6.

**Architecture and loss.** In **A**, we start with a standard flow matching model with a DiT architecture that predicts lattice vectors from mean-pooled atom features. Then, in **B**, we instead treat the lattice as three atom-level tokens and find that it is roughly quality-neutral, but simplifies the architecture to a single token stream and qualitatively stabilizes optimization (Figure 7). Adding the auxiliary

Table 3: Ablation of inference discretization $(t_i)_i$ and training-time distribution $p(t)$ on the validation set. Model **D** from Table 2 corresponds to linear discretization with uniform distribution. Bootstrap means and standard errors over 5000 resamples are reported.

| $(t_i)_i$ | $p(t)$ | Clash Rate | PoseBusters | Vol. Error | EMD PDD |
|---|---|---|---|---|---|
| Log | Uniform | $11.77 \pm 0.37$ | $77.75 \pm 0.44$ | $1.55 \pm 0.04$ | $9.80 \pm 0.03$ |
| | Logit-normal | $6.84 \pm 0.29$ | $31.50 \pm 0.34$ | $1.55 \pm 0.05$ | $9.90 \pm 0.07$ |
| | Beta$(1.8, 1)$ | $11.67 \pm 0.38$ | $82.46 \pm 0.41$ | $1.58 \pm 0.04$ | $9.72 \pm 0.03$ |
| | Ramp | $11.26 \pm 0.37$ | $80.00 \pm 0.42$ | $1.57 \pm 0.04$ | $9.74 \pm 0.03$ |
| Linear | Uniform | $9.79 \pm 0.35$ | $85.23 \pm 0.35$ | $1.55 \pm 0.04$ | $9.66 \pm 0.03$ |
| | Logit-normal | $\mathbf{5.96 \pm 0.27}$ | $75.98 \pm 0.35$ | $\mathbf{1.46 \pm 0.04}$ | $9.56 \pm 0.05$ |
| | Beta$(1.8, 1)$ | $9.56 \pm 0.35$ | $\mathbf{87.34 \pm 0.34}$ | $1.59 \pm 0.04$ | $9.56 \pm 0.03$ |
| | Ramp | $8.91 \pm 0.33$ | $86.08 \pm 0.35$ | $1.53 \pm 0.04$ | $\mathbf{9.55 \pm 0.03}$ |

volume and pairwise distance losses (**C**) and self-conditioning (**D**) results in notable improvements across all metrics.

**Source distribution.** Models **A**–**D** use a fitted lattice prior that closely matches the training distribution (Figure 3), as discussed in Section 3.2. In **E**, we test the standard normal prior and find it to be detrimental, degrading reconstruction metrics to around the level of **B**, with an even larger volume error. The clash rate is the only metric that improves (marginally), though this may be due to the model generating unrealistically large and sparse unit cells, as evidenced by the degraded EMD PDD.

**Timestep settings.** Flow matching is sensitive to both the training-time distribution $p(t)$ and inference discretization $(t_i)_i$. Models for protein and small-molecule generation tend to skew both toward the late-time regime, since it governs local detail [55, 61, 63]. In Table 3, we explore such settings for our crystal domain. Contrary to prior work, we find log discretization consistently harmful, most notably degrading the clash rate and EMD PDD. This supports our hypothesis that crystals are not purely local: global packing geometry is established at intermediate times, so over-emphasizing $t \to 1$ risks starving these stages. Indeed, a logit-normal distribution that focuses mass around the mid-range improves packing metrics over the uniform distribution. However, it also degrades the PoseBusters quality (in fact, collapsing when combined with log discretization), suggesting that underweighting late times can also be problematic. Conversely, a Beta$(1.8, 1)$ [63] distribution shows better PoseBusters validity but performs worse otherwise. Finally, we demonstrate that we can successfully trade off between these endpoints through a *ramp* distribution that places the majority of its mass uniformly on $[0.5, 1]$ (Figure 3a). However, its performance is relatively similar to the beta distribution, so our final models use Beta$(1.8, 1)$ and we leave further exploration to future work.

**Model scale.** Scaling CLARI-M from 88M to CLARI-L at 173M parameters results in a general improvement in reconstruction, which becomes more apparent on the test set. Surprisingly, CLARI-L trains only $\sim$20% slower than CLARI-M, despite having nearly $2\times$ as many parameters. Scaling further to CLARI-H at 350M parameters continues to improve reconstruction metrics, although the clash rate increases.

## 5 Conclusion

We present CLARI, a flow matching model for organic crystal structure prediction that operates directly on a single unit cell using a pair-bias DiT, avoiding both bulk expansion and triangle-update layers. These architectural refinements improve prediction quality while making sampling roughly 15–30$\times$ faster than OXtal on the CSP Blind Tests (5–8$\times$ end-to-end when UMA energy ranking is included). By jointly modelling heavy atoms and hydrogens, CLARI produces structures amenable to direct energy-based ranking, enabling inference-time scaling through best-of-$N$ sampling without decoration or relaxation. We further analyze the sources of these gains through ablations and introduce the CSD Teaching Subset, a benchmark spanning chemically complex systems that are excluded from or underrepresented in prior evaluations.

Our results show that explicit lattice modelling provides an efficient route to generative CSP. Compared with bulk representations, modelling a single unit cell avoids redundancy and yields substantial computational savings, reducing crystal generation from minutes to seconds. This speed unlocks a different use case for CSP: in the case of screening large virtual libraries of millions of candidate

molecules for solid-state properties [23], it may be desirable to simply obtain a *decent* crystal structure within a few-second turnaround time. While prior methods can be accelerated with few-step generation methods [91–95], these techniques are orthogonal to and fully compatible with CLARI's architectural speedups.

At the same time, we view explicit lattice modelling as complementary to bulk approaches such as OXtal. While the unit cell representation is almost always more compact, lattice-free modelling may better accommodate amorphous systems [96], where the unit cell becomes extremely large or ill-defined.

**Limitations and future work.** Search-based CSP offers high accuracy and recall at slow speed, whereas generative CSP offers decent accuracy and recall at fast speed. CLARI's capabilities map onto virtual screening, where high throughput is necessary but accuracy and recall do not need to be watertight, rather than pharmaceutical polymorph screening, where both are necessary and speed is secondary. We do not advocate that CLARI should replace traditional CSP pipelines, but it may supplement them with additional crystal candidates. In preliminary polymorph coverage experiments (Appendix D.2), CLARI-H recovers 61–75% of the observed polymorphs in OXtal test families using only 150 samples per family, though coverage is weaker on the flexible and blind-test families. On the modelling front, CLARI conditions on a 2D molecular graph that is agnostic to stereochemistry, a key limitation for pharmaceutical applications. Conditioning on an explicit 3D conformer or augmenting the graph with chiral tags would allow for more controllable generation. In addition, CLARI requires as input the number of copies of each molecule in the unit cell, which may not always be known at inference time, although it is straightforward to sweep over common values of $Z = 4, 2, 1$. On the data side, we contribute a processed CSD dataset with held-out splits and a new CSD Teaching Subset benchmark, but both leave room for improvement. Constructing test splits that cleanly probe generalization is harder for crystals than for molecules, since graph-based similarity measures are less established for multi-component crystals that RDKit cannot sanitize. Our splits combine refcode grouping with heuristic component-level filtering as a step towards this, but more principled splitting protocols would be valuable for future work. In addition, evaluating generative CSP at scale is hindered by the lack of fast and robust crystal-similarity metrics. We note that Sol is an approximate metric requiring only 8 of 15 molecules to match, which limits the scope of our results. We also observe that CLARI occasionally generates crystals with steric clashes or unphysical voids. Reward alignment or inference-time steering [82, 83, 97] can help reduce these errors. Another promising direction is to apply inference-time steering towards a given powder diffraction pattern [98], space group [50], energy, or density. We hope CLARI encourages a reassessment of architectural choices inherited from biomolecular generative models, and enables virtual screening of solid-state properties for pharmaceuticals, agrochemicals, and organic semiconductors.

**Broader Impacts.** Our work accelerates crystal structure prediction for applications in pharmaceuticals and organic materials. As a dual-use consideration, improved prediction may enable more efficient screening of energetic materials, including the identification of more stable or higher-density explosives. Improved structural understanding may also inform safer formulation and handling practices in industrial uses of energetic materials (e.g., mining and construction). While practical deployment of new energetic materials remains constrained by challenges in synthesis and experimental validation, methods such as CLARI may lower computational barriers to discovery. We therefore emphasize responsible use within established safety and regulatory frameworks.

## Acknowledgments and Disclosure of Funding

The authors thank Matthew Spellings, Michael Kilgour, and Olivier Trottier for helpful discussions. A.H.C. acknowledges the generous support of the Canada 150 Research Chairs program through A.A.-G. L.M. acknowledges support from the Prof. Dr. Sc. Jasna Šimunić-Hrvoić Foundation Fellowship. A.A.-G. thanks Anders G. Frøseth for his generous support. A.A.-G. also acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program. This research is part of the University of Toronto's Acceleration Consortium, which receives funding from the CFREF-2022-00042 Canada First Research Excellence Fund. This research was enabled in part by support provided by SciNet HPC Consortium (https://scinethpc.ca/) and the Digital Research Alliance of Canada (https://www.alliancecan.ca). Computations were performed on the Trillium supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto. This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR0011262E022.

## References

[1] Kenneth Honer, Eren Kalfaoglu, Carlos Pico, Jane McCann, and Jonas Baltrusaitis. Mechanosynthesis of magnesium and calcium salt–urea ionic cocrystal fertilizer materials for improved nitrogen management. *ACS Sustainable Chemistry & Engineering*, 5(10):8546–8550, 2017.

[2] Jingxiang Yang, Chunhua Tony Hu, Xiaolong Zhu, Qiang Zhu, Michael D Ward, and Bart Kahr. DDT polymorphism and the lethality of crystal forms. *Angewandte Chemie*, 129(34): 10299–10303, 2017.

[3] Zhimin Hao and Abul Iqbal. Some aspects of organic pigments. *Chemical Society Reviews*, 26 (3):203–213, 1997.

[4] N Panina, R Van de Ven, P Verwer, Hugo Meekes, E Vlieg, and G Deroover. Polymorph prediction of organic pigments. *Dyes and Pigments*, 79(2):183–192, 2008.

[5] José Miguel Aguilera, Peter J Lillford, and Heribert Watzke. Why food materials science? In *Food materials science: Principles and practice*, pages 3–10. Springer, 2008.

[6] Joseph E Arnold and Graeme M Day. Crystal structure prediction of energetic materials. *Crystal Growth & Design*, 23(8):6149–6160, 2023.

[7] Sarah L Price, Doris E Braun, and Susan M Reutzel-Edens. Can computed crystal energy landscapes help understand pharmaceutical solids? *Chemical Communications*, 52(44): 7065–7077, 2016.

[8] David H Bowskill, Isaac J Sugden, Stefanos Konstantinopoulos, Claire S Adjiman, and Constantinos C Pantelides. Crystal structure prediction methods for organic molecules: State of the art. *Annual Review of Chemical and Biomolecular Engineering*, 12:593–623, 2021.

[9] Stephen R Forrest. The path to ubiquitous and low-cost organic electronic appliances on plastic. *Nature*, 428(6986):911–918, 2004.

[10] Peifu Sun, Dan Liu, Feng Zhu, and Donghang Yan. An efficient solid-solution crystalline organic light-emitting diode with deep-blue emission. *Nature Photonics*, 17(3):264–272, 2023.

[11] Biswajit Bhattacharya, Adam AL Michalchuk, Dorothee Silbernagl, Nobuhiro Yasuda, Torvid Feiler, Heinz Sturm, and Franziska Emmerling. An atomistic mechanism for elasto-plastic bending in molecular crystals. *Chemical Science*, 14(13):3441–3450, 2023.

[12] Hideko Koshima, Kyoko Takechi, Hidetaka Uchimoto, Motoo Shiro, and Daisuke Hashizume. Photomechanical bending of salicylideneaniline crystals. *Chemical Communications*, 47(41): 11423–11425, 2011.

[13] Keith Chadwick, Allan Myerson, and Bernhardt Trout. Polymorphic control by heterogeneous nucleation-a new method for selecting crystalline substrates. *CrystEngComm*, 13(22):6625–6627, 2011.

[14] Dejan-Krešimir Bučar, Graeme M Day, Ivan Halasz, Geoff GZ Zhang, John RG Sander, David G Reid, Leonard R MacGillivray, Melinda J Duer, and William Jones. The curious case of (caffeine)·(benzoic acid): how heteronuclear seeding allowed the formation of an elusive cocrystal. *Chemical Science*, 4(12):4417–4425, 2013.

[15] John Maddox. Crystals from first principles. *Nature*, 335(6187):201–201, 1988.

[16] Lily M Hunnisett, Jonas Nyman, Nicholas Francia, Nathan S Abraham, Claire S Adjiman, Srinivasulu Aitipamula, Tamador Alkhidir, Mubarak Almehairbi, Andrea Anelli, Dylan M Anstine, et al. The seventh blind test of crystal structure prediction: structure generation methods. *Structural Science*, 80(6):517–547, 2024.

[17] Lily M Hunnisett, Nicholas Francia, Jonas Nyman, Nathan S Abraham, Srinivasulu Aitipamula, Tamador Alkhidir, Mubarak Almehairbi, Andrea Anelli, Dylan M Anstine, John E Anthony, et al. The seventh blind test of crystal structure prediction: structure ranking methods. *Structural Science*, 80(6), 2024.

[18] Gregory J. O. Beran. Frontiers of molecular crystal structure prediction for pharmaceuticals and functional organic materials. *Chem. Sci.*, 14:13290–13312, 2023.

[19] Johannes Hoja, Hsin-Yu Ko, Marcus A Neumann, Roberto Car, Robert A DiStasio Jr, and Alexandre Tkatchenko. Reliable and practical computational description of molecular crystal polymorphs. *Science Advances*, 5(1):eaau3338, 2019.

[20] Dong Zhou, Imanuel Bier, Biswajit Santra, Leif D Jacobson, Chuanjie Wu, Adiran Garaizar Suarez, Barbara Ramirez Almaguer, Haoyu Yu, Robert Abel, Richard A Friesner, et al. A robust crystal structure prediction method to support small molecule drug development with large scale validation and blind study. *Nature Communications*, 16(1):2210, 2025.

[21] Anthony M Reilly, Richard I Cooper, Claire S Adjiman, Saswata Bhattacharya, A Daniel Boese, Jan Gerit Brandenburg, Peter J Bygrave, Rita Bylsma, Josh E Campbell, Roberto Car, et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(4): 439–459, 2016.

[22] Majid Mortazavi, Johannes Hoja, Luc Aerts, Luc Quéré, Jacco van de Streek, Marcus A Neumann, and Alexandre Tkatchenko. Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine. *Communications Chemistry*, 2(1):70, 2019.

[23] Hiroyuki Ishii, Shigeaki Obata, Naoyuki Niitsu, Shun Watanabe, Hitoshi Goto, Kenji Hirose, Nobuhiko Kobayashi, Toshihiro Okamoto, and Jun Takeya. Charge mobility calculation of organic semiconductors without use of experimental single-crystal data. *Scientific Reports*, 10 (1):2524, 2020.

[24] Joel Bernstein. *Polymorphism in molecular crystals*, volume 30. Oxford University Press, 2020.

[25] Emily Jin, Andrei Cristian Nica, Mikhail Galkin, Jarrid Rector-Brooks, Kin Long Kelvin Lee, Santiago Miret, Frances H. Arnold, Michael M. Bronstein, Joey Bose, Alexander Tong, and Cheng-Hao Liu. OXtal: An all-atom diffusion model for organic crystal structure prediction. In *The Fourteenth International Conference on Learning Representations*, 2026.

[26] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

[27] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

[28] Brandon M Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John R Kitchin, Daniel S Levine, et al. UMA: A family of universal models for atoms. *arXiv preprint arXiv:2506.23971*, 2025.

[29] Ömer H Omar, Marcos Del Cueto, Tahereh Nematiaram, and Alessandro Troisi. High-throughput virtual screening for organic electronics: a comparative study of alternative strategies. *Journal of Materials Chemistry C*, 9(39):13557–13583, 2021.

[30] Jos PM Lommerse, WD Sam Motherwell, Herman L Ammon, Jack D Dunitz, Angelo Gavezzotti, Detlef WM Hofmann, Frank JJ Leusen, Wijnand TM Mooij, Sarah L Price, Bernd Schweizer, et al. A test of crystal structure prediction of small organic molecules. *Acta Crystallographica Section B: Structural Science*, 56(4):697–714, 2000.

[31] WD Sam Motherwell, Herman L Ammon, Jack D Dunitz, Alexander Dzyabchenko, Peter Erk, Angelo Gavezzotti, Detlef WM Hofmann, Frank JJ Leusen, Jos PM Lommerse, Wijnand TM Mooij, et al. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B: Structural Science*, 58(4):647–661, 2002.

[32] GM Day, WDS Motherwell, HL Ammon, SXM Boerrigter, RAFFAELE GUIDO Della Valle, Elisabetta Venuti, A Dzyabchenko, Jack D Dunitz, Bernd Schweizer, BP Van Eijck, et al. A third blind test of crystal structure prediction. *Acta Crystallographica Section B: Structural Science*, 61(5):511–527, 2005.

[33] Graeme M Day, Timothy G Cooper, Aurora J Cruz-Cabeza, Katarzyna E Hejczyk, Herman L Ammon, Stephan XM Boerrigter, Jeffrey S Tan, Raffaele G Della Valle, Elisabetta Venuti, Jovan Jose, et al. Significant progress in predicting the crystal structures of small organic molecules–a report on the fourth blind test. *Acta Crystallographica Section B: Structural Science*, 65(2):107–125, 2009.

[34] David A Bardwell, Claire S Adjiman, Yelena A Arnautova, Ekaterina Bartashevich, Stephan XM Boerrigter, Doris E Braun, Aurora J Cruz-Cabeza, Graeme M Day, Raffaele G Della Valle, Gautam R Desiraju, et al. Towards crystal structure prediction of complex organic compounds–a report on the fifth blind test. *Acta Crystallographica Section B: Structural Science*, 67(6):535–551, 2011.

[35] Tzu-Jen Lin, Cheng-Rong Hsing, Ching-Ming Wei, and Jer-Lai Kuo. Structure prediction of the solid forms of methanol: an ab initio random structure searching approach. *Physical chemistry chemical Physics*, 18(4):2736–2746, 2016.

[36] David H Case, Josh E Campbell, Peter J Bygrave, and Graeme M Day. Convergence properties of crystal structure prediction by quasi-random sampling. *Journal of chemical theory and computation*, 12(2):910–924, 2016.

[37] Charles Richard Arthur Catlow, John Meurig Thomas, M Freeman, Clive, A Wright, Paul, and G Bell, Robert. Simulating and predicting crystal structures. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 442(1914):85–96, 1993.

[38] Farren Curtis, Timothy Rose, and Noa Marom. Evolutionary niching in the GAtor genetic algorithm for molecular crystal structure prediction. *Faraday discussions*, 211:61–77, 2018.

[39] Vahe Gharakhanyan, Yi Yang, Luis Barroso-Luque, Muhammed Shuaibi, Daniel S Levine, Kyle Michel, Viachaslau Bernat, Misko Dzamba, Xiang Fu, Meng Gao, et al. FastCSP: Accelerated molecular crystal structure prediction with universal model for atoms. *arXiv preprint arXiv:2508.02641*, 2025.

[40] Akshay Subramanian, Elton Pan, Juno Nam, Maurice Weiler, Shuhui Qu, Cheol Woo Park, Tommi S Jaakkola, Elsa Olivetti, and Rafael Gomez-Bombarelli. PackFlow: Generative molecular crystal structure prediction via reinforcement learning alignment. *arXiv preprint arXiv:2602.20140*, 2026.

[41] Cheng Zeng, Harry W Sullivan, Thomas Egg, Maya M Martirossyan, Philipp Höllmer, Jirui Jin, Richard G Hennig, Adrian Roitberg, Stefano Martiniani, Ellad B Tadmor, et al. MolCrystalFlow: Molecular crystal structure prediction via flow matching. *arXiv preprint arXiv:2602.16020*, 2026.

[42] 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.

[43] Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. *Advances in Neural Information Processing Systems*, 36:17464–17497, 2023.

[44] Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. In *The Twelfth International Conference on Learning Representations*, 2024.

[45] 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.

[46] 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.

[47] 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.

[48] Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Xuechen Shao, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. CrystalFlow: a flow-based generative model for crystalline materials. *Nature Communications*, 16(1):9267, 2025.

[49] 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.

[50] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[51] Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi S. Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *ICML*, pages 40001–40039, 2023.

[52] Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations*, 2024.

[53] Bowen Jing, Ezra Erives, Peter Pao-Huang, Gabriele Corso, Bonnie Berger, and Tommi Jaakkola. EigenFold: Generative protein structure prediction with diffusion models. *arXiv preprint arXiv:2304.02198*, 2023.

[54] Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold meets flow matching for generating protein ensembles. In *Forty-first International Conference on Machine Learning*, 2024.

[55] Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Proteina: Scaling flow-based protein structure generative models. In *The Thirteenth International Conference on Learning Representations*, 2025.

[56] Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-Proteina: Atomistic protein generation via partially latent flow matching. In *The Fourteenth International Conference on Learning Representations*, 2026.

[57] Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026.

[58] Hannes Stark, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Harmonic self-conditioned flow matching for joint multi-ligand docking and binding site design. In *Forty-first International Conference on Machine Learning*, 2024.

[59] Yuyang Wang, Ahmed A. A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: Simplified scalable conformer generation. In *Forty-first International Conference on Machine Learning*, 2024.

[60] Ian Dunn and David R Koes. FlowMol3: flow matching for 3D de novo small-molecule generation. *Digital Discovery*, 2026.

[61] Ross Irwin, Alessandro Tibo, Jon Paul Janet, and Simon Olsson. SemlaFlow – efficient 3D molecular generation with latent attention and equivariant flow matching. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

[62] Danny Reidenbach, Filipp Nikitin, Olexandr Isayev, and Saee Gopal Paliwal. Applications of modular co-design for de novo 3d molecule generation. *Digital Discovery*, 5(2):754–768, 2026.

[63] Carlos Vonessen, Charles Harris, Miruna Cretu, and Pietro Liò. TABASCO: A fast, simplified model for molecular generation with improved physical quality. *arXiv preprint arXiv:2507.00899*, 2025.

[64] Luis M. Antunes, Keith T. Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling. *Nature Communications*, 15(1):10570, Dec 2024. doi: 10.1038/s41467-024-54639-7.

[65] Zihao Li, Zhichen Zeng, Xiao Lin, Feihao Fang, Yanru Qu, Zhe Xu, Zhining Liu, Xuying Ning, Tianxin Wei, Ge Liu, Hanghang Tong, and Jingrui He. Flow matching meets biology and life science: A survey. *arXiv preprint arXiv:2507.17731*, 2026.

[66] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[67] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

[68] Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

[69] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024.

[70] Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. In *The Eleventh International Conference on Learning Representations*, 2023.

[71] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.

[72] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[73] Zihan Qiu, Zekun Wang, Bo Zheng, Zeyu Huang, Kaiyue Wen, Songlin Yang, Rui Men, Le Yu, Fei Huang, Suozhi Huang, Dayiheng Liu, Jingren Zhou, and Junyang Lin. Gated attention for large language models: Non-linearity, sparsity, and attention-sink-free. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[74] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240): 1–113, 2023.

[75] Noam Shazeer. GLU variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020.

[76] Alexander Tong, Kilian FATRAS, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. Expert Certification.

[77] Leon Klein, Andreas Krämer, and Frank Noe. Equivariant flow matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[78] Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, et al. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv*, pages 2024–11, 2025.

[79] Harold W Kuhn. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[80] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Foundations of Crystallography*, 32(5):922–923, 1976.

[81] Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, et al. Inference-time scaling for diffusion models beyond scaling denoising steps. *arXiv preprint arXiv:2501.09732*, 2025.

[82] Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Forty-second International Conference on Machine Learning*, 2025.

[83] Marta Skreta, Tara Akhound-Sadegh, Viktor Ohanesian, Roberto Bondesan, Alan Aspuru-Guzik, Arnaud Doucet, Rob Brekelmans, Alexander Tong, and Kirill Neklyudov. Feynman-kac correctors in diffusion: Annealing, guidance, and product of experts. In *Forty-second International Conference on Machine Learning*, 2025.

[84] Colin R. Groom, Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The Cambridge Structural Database. *Acta Crystallographica Section B*, 72(2):171–179, Apr 2016.

[85] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35: 26565–26577, 2022.

[86] R Alex Mayo, Alberto Otero-de-la Roza, and Erin R Johnson. Development and assessment of an improved powder-diffraction-based method for molecular crystal structure similarity. *CrystEngComm*, 24(47):8326–8338, 2022.

[87] James Alexander Chisholm and Sam Motherwell. COMPACK: A program for identifying crystal structure similarity using distances. *Applied Crystallography*, 38(1):228–231, 2005.

[88] Martin Buttenschoen, Garrett M Morris, and Charlotte M Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.

[89] Daniel Widdowson and Vitaliy Kurlin. Resolving the data ambiguity for periodic crystals. *Advances in Neural Information Processing Systems*, 35:24625–24638, 2022.

[90] Gary M Battle, Gregory M Ferrence, and Frank H Allen. Applications of the cambridge structural database in chemical education. *Applied Crystallography*, 43(5):1208–1223, 2010.

[91] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *ICML*, pages 32211–32252, 2023.

[92] Nicholas Matthew Boffi, Michael Samuel Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[93] Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[94] Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026.

[95] Mingyang Deng, He Li, Tianhong Li, Yilun Du, and Kaiming He. Generative modeling via drifting. *arXiv preprint arXiv:2602.04770*, 2026.

[96] Manuel Cordova, Martins Balodis, Albert Hofstetter, Federico Paruzzo, Sten O Nilsson Lill, Emma SE Eriksson, Pierrick Berruyer, Bruno Simões de Almeida, Michael J Quayle, Stefan T Norberg, et al. Structure determination of an amorphous drug through large-scale nmr predictions. *Nature Communications*, 12(1):2964, 2021.

[97] Peter Potaptchik, Adhi Saravanan, Abbas Mammadov, Alvaro Prat, Michael S Albergo, and Yee Whye Teh. Meta flow maps enable scalable reward alignment. *arXiv preprint arXiv:2601.14430*, 2026.

[98] Qi Li, Rui Jiao, Liming Wu, Tiannian Zhu, Wenbing Huang, Shifeng Jin, Yang Liu, Hongming Weng, and Xiaolong Chen. Powder diffraction crystal structure determination using generative models. *Nature Communications*, 16(1):7428, 2025.

[99] Beatriz Cordero, Verónica Gómez, Ana E. Platero-Prats, Marc Revés, Jorge Echeverría, Eduard Cremades, Flavia Barragán, and Santiago Alvarez. Covalent radii revisited. *Dalton Trans.*, pages 2832–2838, 2008.

[100] Cambridge Crystallographic Data Centre. CSD elemental radii, August 2015. URL https://www.ccdc.cam.ac.uk/media/Elemental_Radii_Alvarez.xlsx.

[101] Keller Jordan, Yuchen Jin, Vlado Boza, You Jiacheng, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024. URL https://kellerjordan.github.io/posts/muon/.

[102] Jingyuan Liu, Jianlin Su, Xingcheng Yao, Zhejun Jiang, Guokun Lai, Yulun Du, Yidao Qin, Weixin Xu, Enzhe Lu, Junjie Yan, Yanru Chen, Huabin Zheng, Yibo Liu, Shaowei Liu, Bohong Yin, Weiran He, Han Zhu, Yuzhi Wang, Jianzhou Wang, Mengnan Dong, Zheng Zhang, Yongsheng Kang, Hao Zhang, Xinran Xu, Yutao Zhang, Yuxin Wu, Xinyu Zhou, and Zhilin Yang. Muon is scalable for LLM training. *arXiv preprint arXiv:2502.16982*, 2025.

[103] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[104] Yi Yang, Rithwik Tom, Jose AGL Wui, Jonathan E Moussa, and Noa Marom. Genarris 3.0: Generating close-packed molecular crystal structures with rigid press. *Journal of Chemical Theory and Computation*, 21(21):11318–11332, 2025.

# A Dataset

## A.1 Crystal processing pipeline

The Cambridge Structural Database (CSD) requires a license. We obtained an academic license via the University of Toronto.

We download CSD entry metadata using the CSD Python API and the raw CIF and MOL2 files using ConQuest. Given a CSD entry, the following workflow is performed:

1. Metadata filtering according to the criteria discussed in Section 4.1.
2. Parsing the crystal's asymmetric unit from its CIF and MOL2 file.
3. Caching component isomorphisms within the asymmetric unit.
4. Expanding the asymmetric unit to the full unit cell by applying space group operations.
5. Distance-based filtering to remove polymers and steric clashes.
6. Unit cell featurization (e.g., SMILES for splitting on test components).

Crystals that are successfully processed without error are aggregated and split into training, validation, and test datasets. We elaborate on the critical steps 3 and 5.

**Isomorphism caching.** For crystal alignment (Section 3.4), it is useful to know which components in the unit cell are isomorphic, and if so, what those isomorphisms are. We cache these during our data processing pipeline to avoid expensive isomorphism checks during training. To further improve efficiency, we compute these on the asymmetric unit, which generally has many fewer components than the full unit cell. When the asymmetric unit is replicated in step (4), these isomorphisms can then be easily extended. Note that we do not attempt to intractably enumerate all isomorphisms between components but rather find a single one, if it exists.

**Distance-based filtering.** We more strictly filter out polymeric or noisy crystals (e.g., due to improper disorder specification) by checking that all atoms are sufficiently distanced. Let $M_1$ and $M_2$ be two (not necessarily distinct) components obtained after step (4). Let $\boldsymbol{p}_1$ be the position of an atom $a_1$ from $M_1$, and similarly for $\boldsymbol{p}_2$ and $a_2$ from $M_2$. We check that

$$d_{\boldsymbol{L}}(\boldsymbol{p}_1, \boldsymbol{p}_2) \geq \alpha \left( r(a_1) + r(a_2) \right), \quad \text{where } \alpha = \begin{cases} 0.6, & \text{if } a_1 \text{ or } a_2 \text{ is a metal,} \\ 0.6, & \text{if } a_1 \text{ and } a_2 \text{ are bonded,} \\ 1.0, & \text{otherwise,} \end{cases} \tag{2}$$

where $\boldsymbol{L}$ is the cell lattice and $r(\cdot)$ is the CSD covalent radius (Å) [99, 100]. Threshold distances $\alpha$ were chosen based on the periodic distance distribution of CSD (Figure 6). A minor technicality is that two distinct but exactly superimposable components can arise, for example, when a component in the asymmetric unit is invariant under the space group. In this case, we should discard one component rather than rejecting the entire CSD entry. If $M_1 \neq M_2$ are isomorphic but clashing, we can check for this case by performing an optimal atomic assignment [79] and checking that all atoms are within $0.01$ Å deviation.

## A.2 Dataset splitting

The OXtal test crystal families and the CSD Teaching Subset are combined to create our held-out test pool. To mitigate leakage, we exclude entire 6-letter refcode families of test entries and additionally exclude crystals sharing an RDKit-sanitizable component with any test structure. The component must be RDKit sanitizable so we can perform the equality check using its canonical SMILES rather than expensive graph isomorphism checks. Moreover, the component must also have over 7 heavy atoms, so that ubiquitous but small molecules such as water or hexafluorophosphate are not considered. The validation set is created by holding out 1000 refcode families from the remainder. The final split counts, before benchmark-specific evaluation filters, contain 917,014 training, 1048 validation, and 2996 test examples across 859,866, 1000, and 919 families, respectively; the test count includes all held-out crystals, including those with more than 512 atoms in the unit cell.

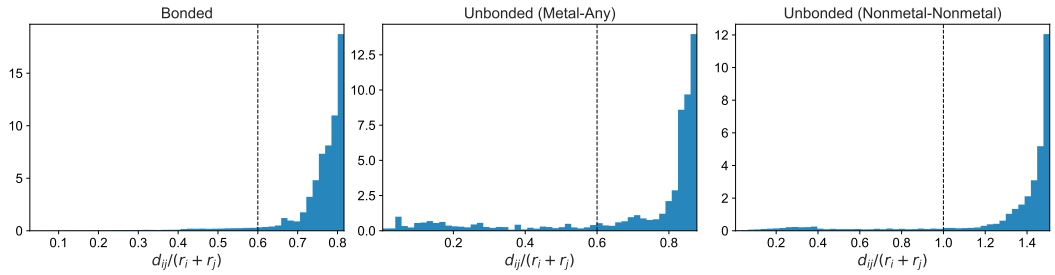



Figure 6: Distribution of normalized periodic interatomic distances $d_{\boldsymbol{L}}(\boldsymbol{p}_1, \boldsymbol{p}_2)/(r(a_1) + r(a_2))$ across the CSD, with the threshold $\alpha$ (Equation 2) shown in a dashed line. For each atom in the CSD, we gather distances of adjacent bonded atoms (left) and the nearest unbonded atoms. The latter case is further broken down based on whether one of the atoms is a metal (middle) or both are nonmetals (right). Histograms are each truncated at their lower $0.5\%$ tail.

# B Modelling

## B.1 Source distribution

We decompose lattice matrices $\boldsymbol{L} = (\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3)^\top \in \mathbb{R}^{3\times 3}$ of an $N$-atom unit cell into three components and model each as independent Gaussians fitted to the CSD training set:

- The atom density $\rho \sim \mathcal{N}(\mu_\rho, \sigma_\rho^2)$, where $\rho = N/V$ for $V = |\det \boldsymbol{L}|$.
- The cell angles $\alpha, \beta, \gamma \sim \mathcal{N}(\mu_\circ, \sigma_\circ^2)$ i.i.d., where $\alpha = \text{angle}(\boldsymbol{l}_1, \boldsymbol{l}_2)$ and similarly for $\beta, \gamma$.
- The normalized cell lengths $(a, b, c) \sim \mathcal{N}(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell)$, where $a = ||\boldsymbol{l}_1||_2/V^{1/3}$ and $b, c$ are defined similarly, sorted such that $a \le b \le c$. Note that $\boldsymbol{\Sigma}_\ell$ is not diagonal.

Given a sample $(\rho, \alpha, \beta, \gamma, a, b, c)$, we can reconstruct a matrix $\boldsymbol{L}$ that adheres to the given parameters. To randomize the pose of the matrix, we finally apply a random signed permutation and rotation.

## B.2 Architecture

Let $\text{Linear}(\cdot)$ and $\text{LN}(\cdot)$ denote a linear layer and LayerNorm. Let

$$\text{Mod}(\boldsymbol{x}, \boldsymbol{c}) = \text{Linear}(\boldsymbol{c}) \odot \boldsymbol{x} + \text{Linear}(\boldsymbol{c}), \tag{3}$$

$$\text{Gate}(\boldsymbol{x}, \boldsymbol{c}) = \text{Linear}(\boldsymbol{c}) \odot \boldsymbol{x}, \tag{4}$$

$$\text{AdaLN}(\boldsymbol{x}, \boldsymbol{c}) = \text{Mod}(\text{LN}(\boldsymbol{x}), \boldsymbol{c}), \tag{5}$$

$$\text{Transition}(\boldsymbol{x}) = \text{Linear}(\text{SwiGLU}(\text{Linear}(\boldsymbol{x}))), \tag{6}$$

and let $\text{PairAttention}(\boldsymbol{h}, \boldsymbol{z})$ denote attention with pairwise bias via $\text{Linear}(\boldsymbol{z})$. The architecture of CLARI is summarized in Algorithm 1. The input features to CLARI are listed in Table 4.

## B.3 Inference batching

The dataloader yields one crystal per batch; we then replicate it $B(n)$ times in the batch dimension for parallel sampling, where $B(n) = 1000, 500, 200, 25, 1$ for $n < 200, < 300, < 500, \le 1000, > 1000$ atoms. On an out-of-memory error, $B$ is halved and the chunk is retried. Before beginning timing, we warm up compilation with dummy runs on the first $5$ batches.

## B.4 Auxiliary losses

The volume and pair losses are applied between the one-step estimate $\hat{\boldsymbol{x}}_1 = \boldsymbol{x}_t + (1-t)\, v_\theta(\boldsymbol{x}_t, t)$ and ground truth $\boldsymbol{x}_1$. The auxiliary volume loss is the relative error of the predicted volume:

$$\mathcal{L}_{\text{vol}} = \left| \frac{|\det \hat{\boldsymbol{L}}_1|}{|\det \boldsymbol{L}_1|} - 1 \right|, \tag{7}$$

where $\hat{\boldsymbol{L}}_1$ and $\boldsymbol{L}_1$ are the lattice matrices obtained from $\hat{\boldsymbol{x}}_1$ and $\boldsymbol{x}_1$, respectively. Now, let $\hat{d}_{ij}$ and $d_{ij}$ be the periodic distance (Å) between the $i$-th and $j$-th atoms in $\hat{\boldsymbol{x}}_1$ and $\boldsymbol{x}_1$, respectively, and $\alpha_{ij}$ be

**Algorithm 1** CLARI architecture.

**Require:** Sequence dimension $d$, pair dimension $d_z$, conditioning dimension $d_c$, trunk depth $D$, number of attention heads $H$ (Table 5).

Create conditioning features

$\boldsymbol{c} \leftarrow$ embed conditioning features (Table 4) ▷ $\boldsymbol{c} \in \mathbb{R}^{d_c}$

**for** $i = 1, 2$ **do**

  $\boldsymbol{c} \leftarrow \boldsymbol{c} + \mathrm{Transition}(\mathrm{LN}(\boldsymbol{c}))$

$\boldsymbol{c} \leftarrow \mathrm{LN}(\boldsymbol{c})$

Create sequence features

$\boldsymbol{h} \leftarrow$ embed sequence features (Table 4) ▷ $\boldsymbol{h} \in \mathbb{R}^{n \times d}$

$\boldsymbol{h} \leftarrow \mathrm{Mod}(\tanh(\boldsymbol{h}), \boldsymbol{c})$

$(\boldsymbol{h}, \boldsymbol{u}, \boldsymbol{v}) \leftarrow \mathrm{Transition}(\boldsymbol{h})$ ▷ split as $d \oplus d_z \oplus d_z$

Create pair features

$\boldsymbol{z} \leftarrow$ embed pair features (Table 4) ▷ $\boldsymbol{z} \in \mathbb{R}^{n \times n \times d_z}$

$\boldsymbol{z} \leftarrow \boldsymbol{z} + (\boldsymbol{u}_i + \boldsymbol{v}_j)_{ij}$

$\boldsymbol{z} \leftarrow \mathrm{Mod}(\tanh(\boldsymbol{z}), \boldsymbol{c})$

Main trunk

**for** $\ell = 1, \ldots, D$ **do**

  $\boldsymbol{h} \leftarrow \boldsymbol{h} + \mathrm{Gate}(\mathrm{PairAttention}(\mathrm{AdaLN}(\boldsymbol{h}, \boldsymbol{c}), \boldsymbol{z}), \boldsymbol{c})$

  $\boldsymbol{h} \leftarrow \boldsymbol{h} + \mathrm{Gate}(\mathrm{Transition}(\mathrm{AdaLN}(\boldsymbol{h}, \boldsymbol{c})), \boldsymbol{c})$

$\boldsymbol{v} \leftarrow \mathrm{Transition}(\mathrm{AdaLN}(\boldsymbol{h}, \boldsymbol{c}))$ ▷ $\boldsymbol{v} \in \mathbb{R}^{n \times 3}$

**return** $\boldsymbol{v}$

Table 4: Input features consumed by CLARI. Within each stream, we take the sum of the individual embeddings as the final sequence, pair, or conditioning features.

| | Input | Featurization |
|---|---|---|
| Seq. | Cartesian $\boldsymbol{x}_t$ | Linear and sinusoidal. |
| | Fractional $\boldsymbol{x}_t$ | Linear and sinusoidal (periodic in 1). |
| | Self-cond. $\hat{\boldsymbol{x}}_1$ | Linear and sinusoidal. |
| | Element | Embedding. Uncommon elements map to a $*$ type. Lattice nodes are treated as 3 new elements. |
| | Atomic charges | Linear and embedding with bins $\{\text{-2-}, \text{-1}, 0, 1, \text{2+}\}$. |
| | Atomic degrees | Linear and embedding with bins $\{0, \ldots, \text{9+}\}$. |
| | Atomic radii | Linear of covalent and VDW radii. |
| | Adjacent bonds | Linear projection of binary indicator vector. |
| Pair | Bonds | Embedding. |
| | Topological dist. | Embedding with bins $\{0, \ldots, 15, \text{16+}, \infty\}$. |
| | Cartesian dist. | Embedding with 128 bins over $[0, 32]$ Å. |
| | Periodic dist. | Embedding with 128 bins over $[0, 32]$ Å. |
| Cond. | Timestep $t$ | Sinusoidal. |
| | Lattice $\boldsymbol{L}$ | Linear projection of $\boldsymbol{L}$, $\boldsymbol{L}^{-1}$, $\boldsymbol{L}^\top \boldsymbol{L}$, and $\det \boldsymbol{L}$. |
| | Formula | Linear projection of count vector. |
| | Self-cond. | Embedding (binary) for whether $\hat{\boldsymbol{x}}_1 = \varnothing$. |

the distance threshold from Equation 2. The auxiliary pair loss is

$$\mathcal{L}_{\text{pair}} = \sum_{(i,j) \in \Lambda} \left[ \left| \hat{d}_{ij} - d_{ij} \right| + 5 \cdot \max\left(0, \alpha_{ij} - \hat{d}_{ij}\right) \right], \tag{8}$$

where $\Lambda = \{i \neq j \mid d_{ij} < 15\,\text{Å} \text{ or } \hat{d}_{ij} < \alpha_{ij}\}$. The pair loss combines a distance matching term with a clash term. To compute the periodic distances tractably and stably, we use the estimate:

$$d_{\boldsymbol{L}}(\boldsymbol{p}_i, \boldsymbol{p}_j) \leq \min_{\boldsymbol{z} \in \{-1,0,1\}^3} \left\| \boldsymbol{p}_i - \boldsymbol{p}_j - \boldsymbol{L}^\top \boldsymbol{z} \right\|_2,$$

which tends to be exact for most crystals.

### B.5 Optimal-transport coupling

We approximately align $\boldsymbol{x}_1$ to $\boldsymbol{x}_0$ before constructing the interpolant $\boldsymbol{x}_t$ by composing three tractable substeps. First, we align the pair purely on their lattices to obtain a good initial pose. Then, we iterate between permutation and pose alignment twice. Let $(\boldsymbol{L}_1, \boldsymbol{C}_1)$ and $(\boldsymbol{L}_0, \boldsymbol{C}_0)$ be the lattice and atomic rows of $\boldsymbol{x}_1$ and $\boldsymbol{x}_0$, respectively. The steps are now described:

**Exact lattice alignment.** The goal is to find a signed permutation $\boldsymbol{\Pi}$ and rotation $\boldsymbol{R}$ that minimize $||\boldsymbol{\Pi}\boldsymbol{L}_1\boldsymbol{R}^\top - \boldsymbol{L}_0||_F$. We can exactly solve this joint problem by enumerating over all such $\boldsymbol{\Pi}$, and for each, finding the optimal $\boldsymbol{R}$ using the Kabsch algorithm [80].

**Permutation alignment.** With pose fixed, the goal is to permute the rows of $\boldsymbol{x}_1$ to minimize its distance to $\boldsymbol{x}_0$. The lattice and atomic rows can be handled independently. For the former, we can again brute-force search across all signed permutations. For the latter, we need to find a permutation $\boldsymbol{\Pi}$ that minimizes $||\boldsymbol{\Pi}\boldsymbol{C}_1 - \boldsymbol{C}_0||_F$ such that $\boldsymbol{\Pi}$ is also an automorphism of the crystal molecular graph. Since enumerating all such automorphisms is intractable, we use the isomorphisms cached from our dataset processing (Appendix A) to get an approximate answer. Specifically, for all isomorphic components between $\boldsymbol{C}_1$ and $\boldsymbol{C}_0$, we compute their RMSD using the atom assignment given by the cached isomorphism. Then, we resolve the assignment of bodies in $\boldsymbol{x}_1$ to those in $\boldsymbol{x}_0$ using the Hungarian algorithm [79]. Note that this does not search the full space because we only consider one of many possible isomorphisms between components.

**Pose alignment.** With order fixed, we can $\mathrm{SO}(3)$-align $\boldsymbol{x}_1$ to $\boldsymbol{x}_0$ using a weighted Kabsch algorithm in which the three lattice rows carry equal weight to the $N$ atom rows.

### B.6 Hyperparameters

Table 5 lists the architectural and training hyperparameters for the medium, large, and huge CLARI models. We use a feedforward expansion of $2.66\times$ to match the parameters of a standard $4\times$ non-GLU feedforward, except for CLARI-H, which uses $4\times$. We train the Transformer weight matrices using Muon [101, 102] and all other parameters using Adam [103]. We also maintain an exponential moving average (EMA) of all parameters for sampling and evaluation.

Table 5: Training and architectural hyperparameters for CLARI.

| Parameters | CLARI-M 88M | CLARI-L 173M | CLARI-H 350M |
|---|---|---|---|
| Sequence dim $d$ | 512 | 768 | 1024 |
| Pair dim $d_z$ | 64 | 64 | 64 |
| Conditioning dim $d_c$ | 512 | 512 | 512 |
| Depth $D$ | 16 | 16 | 16 |
| Heads $H$ | 8 | 12 | 16 |
| Feedforward expansion | $2.66\times$ | $2.66\times$ | $4\times$ |
| Training steps | 150,000 | 150,000 | 150,000 |
| Effective batch size | 128 | 256 | 512 |
| Optimizer | Muon, Adam | Muon, Adam | Muon, Adam |
| Learning rate | 0.0005 | 0.0005 | 0.0005 |
| Warmup steps | 5000 | 5000 | 5000 |
| Weight decay | 0 | 0 | 0 |
| EMA decay | 0.999 | 0.999 | 0.999 |
| GPUs | $4\times$ H100 | $8\times$ H100 | $16\times$ H100 |
| Training time | 14 h | 17 h | 21 h |

## C Metrics

We consider *quality* metrics, assessing the inherent plausibility of a single sample, and *reconstruction* metrics, measuring agreement with a reference. Clash rate and PoseBusters are quality metrics while the others are reconstruction metrics.

**Clash rate.** We consider two *distinct* components within a crystal clashing if an atom from each has periodic distance less than their sum of covalent radii (i.e., Equation 2 with $\alpha = 1$). We report the fraction of crystals with at least one pair of clashing components.

**PoseBusters.** We report the average PoseBusters [88] validity of all components in the unit cell, following the settings from Vonessen et al. [63]. Since PoseBusters relies on RDKit, we process the predicted crystal by disconnecting organometallic bonds and then excluding any components that (i) have no bonds, (ii) contain an element outside of H, C, N, O, F, P, S, Cl, Br, I, (iii) are not RDKit sanitizable, (iv) contain added hydrogens after sanitization. That is, PoseBusters is run only on a subset of applicable components. This metric is averaged over all crystals that have at least one applicable component.

**Relative volume error.** This is $\mathcal{L}_{\text{vol}}$ in Equation 7.

**EMD PDD.** The pointwise distance distribution (PDD) [89] is an isometry invariant of periodic point sets, computed from sorted pairwise interatomic distances of the nearest $k = 100$ neighbours to each point. We report the ($L_1$) earth mover's distance between the PDDs of the predicted and reference crystals.

**COMPACK RMSD$_{15}$.** COMPACK [87] matches a 15-molecule cluster between the predicted and reference packings, ignoring hydrogens and all bond information so that the comparison is sensitive only to heavy-atom packing. For fair comparison we inherit the OXtal configuration verbatim: $50\%$ distance tolerance and $75^\circ$ angle tolerance. These tolerances are loose, but we keep them unchanged so that all reported numbers are directly comparable to OXtal. We also only consider RMSD$_{15} < 2.0$ Å. Samples clash if any inter-body atom pair sits at a distance less than the sum of their van der Waals radii minus $0.7$ Å. A target is counted as solved (Sol) if at least one of the $k$ generated candidates obtains a COMPACK match rate of at least 8 out of 15 molecules to the experimental structure with RMSD$_{15} < 2.0$ Å and no detected collisions. OXtal additionally defines collision, packing-similarity, and conformer-recovery metrics. We focus on Sol since it is the critical metric for downstream applications, and report the extended metrics in Table 8. COMPACK is computationally expensive and is therefore reserved for the test set.

**OXtal comparison protocol.** For direct comparison we adopt the OXtal test set of 119 crystals: 50 rigid, 50 flexible, and the CSP blind-test entries (6 from CSP5, 5 from CSP6, and 8 from CSP7). Model inputs are taken from the CSD entry with the lowest R-factor, i.e., molecular graphs and multiplicities. When a crystal is associated with multiple CSD refcodes of the same family, we evaluate against the three entries with the lowest R-factors and report the best COMPACK match. For the Teaching Subset, we only match against the directly corresponding refcode. Generated structures are evaluated as produced by the model, with no relaxation, energy minimization, or DFT polish. The only post-processing is the UMA-based ranking described in Section 3.5.

# D Additional results

## D.1 Extended baseline comparisons

Tables 6 and 7 report the Sol@$k$ results of Table 1 with standard errors under the strict $15/15$ and $8/15$ molecule match requirements, respectively. Table 8 reports the collision (Col$_S$), packing similarity (Pac$_S$, Pac$_C$), and conformer recovery (Rec$_S$, Rec$_C$) metrics of Jin et al. [25], where $S$ and $C$ denote sample-level and crystal-level aggregation. CLARI shows no collisions and outperforms OXtal on conformer recovery. CLARI appears to underperform on packing similarity, which we believe is due to differences between bulk and unit cell methods: a unit cell repetitively tiles space and so offers fewer packing environments to match against, whereas a bulk crop gives OXtal samples more chances to find matching clusters, even if they ultimately do not pass the RMSD threshold needed for Sol.

Table 6: The Sol@$k$ metrics from Table 1, except with a stricter 15-molecule match requirement. Bootstrap means and standard errors over 5000 resamples are reported, where appropriate.

| Method | $n_s$ | $k$ | Rigid (50) | Flexible (50) | Teach. (773) |
|---|---|---|---|---|---|
| CLARI-M | 30 | 30 | 0.288 ± 0.040 | 0.037 ± 0.024 | 0.140 ± 0.008 |
| CLARI-M | 150 | 30 | 0.344 ± 0.043 | 0.045 ± 0.025 | 0.168 ± 0.008 |
| CLARI-L | 30 | 30 | 0.373 ± 0.043 | 0.071 ± 0.031 | 0.188 ± 0.009 |
| CLARI-L | 150 | 30 | 0.435 ± 0.045 | 0.131 ± 0.036 | 0.223 ± 0.009 |
| CLARI-H | 30 | 30 | 0.461 ± 0.044 | 0.131 ± 0.037 | 0.292 ± 0.009 |
| CLARI-H | 150 | 30 | **0.539 ± 0.039** | **0.208 ± 0.039** | **0.345 ± 0.009** |
| CLARI-M | 400 | 200 | 0.581 ± 0.040 | 0.189 ± 0.037 | 0.316 ± 0.008 |
| CLARI-L | 400 | 200 | 0.684 ± 0.038 | 0.294 ± 0.039 | 0.389 ± 0.008 |
| CLARI-H | 400 | 200 | **0.724 ± 0.028** | **0.396 ± 0.036** | **0.518 ± 0.009** |
| CLARI-M | 1000 | 1000 | 0.720 | 0.380 | 0.446 |
| CLARI-L | 1000 | 1000 | **0.840** | 0.460 | 0.524 |
| CLARI-H | 1000 | 1000 | 0.800 | **0.520** | **0.655** |

| Method | $n_s$ | $k$ | CSP5 (6) | CSP6 (5) | CSP7 (8) |
|---|---|---|---|---|---|
| CLARI-M | 30 | 30 | 0.055 ± 0.090 | 0.007 ± 0.036 | 0.101 ± 0.060 |
| CLARI-M | 150 | 30 | 0.065 ± 0.092 | 0.027 ± 0.069 | 0.106 ± 0.045 |
| CLARI-L | 30 | 30 | 0.187 ± 0.139 | 0.070 ± 0.109 | 0.121 ± 0.021 |
| CLARI-L | 150 | 30 | 0.305 ± 0.159 | **0.155 ± 0.136** | 0.122 ± 0.020 |
| CLARI-H | 30 | 30 | 0.333 ± 0.141 | 0.075 ± 0.110 | 0.132 ± 0.032 |
| CLARI-H | 150 | 30 | **0.480 ± 0.141** | 0.141 ± 0.132 | **0.143 ± 0.044** |
| CLARI-M | 400 | 200 | 0.331 ± 0.129 | 0.066 ± 0.094 | 0.145 ± 0.046 |
| CLARI-L | 400 | 200 | 0.602 ± 0.154 | **0.399 ± 0.139** | 0.125 ± 0.000 |
| CLARI-H | 400 | 200 | **0.748 ± 0.093** | 0.300 ± 0.108 | **0.205 ± 0.080** |
| CLARI-M | 1000 | 1000 | 0.500 | 0.200 | 0.250 |
| CLARI-L | 1000 | 1000 | **0.833** | **0.600** | 0.125 |
| CLARI-H | 1000 | 1000 | **0.833** | **0.600** | **0.375** |

Table 7: The Sol@$k$ metrics from Table 1. Bootstrap means and standard errors over 5000 resamples are reported, where appropriate.

| Method | $n_s$ | $k$ | Rigid (50) | Flexible (50) | Teach. (773) |
|---|---|---|---|---|---|
| $\text{DFT}_{\text{avg}}$ | – | – | – | – | – |
| OXtal | 30 | 30 | 0.300 | 0.220 | – |
| CLARI-M | 30 | 30 | $0.697 \pm 0.039$ | $0.241 \pm 0.044$ | $0.442 \pm 0.010$ |
| CLARI-M | 150 | 30 | $0.712 \pm 0.035$ | $0.260 \pm 0.043$ | $0.470 \pm 0.010$ |
| CLARI-L | 30 | 30 | $0.731 \pm 0.040$ | $0.287 \pm 0.044$ | $0.461 \pm 0.010$ |
| CLARI-L | 150 | 30 | $0.772 \pm 0.039$ | $0.346 \pm 0.047$ | $0.484 \pm 0.010$ |
| CLARI-H | 30 | 30 | $0.766 \pm 0.034$ | $0.378 \pm 0.044$ | $0.584 \pm 0.010$ |
| CLARI-H | 150 | 30 | $\mathbf{0.798 \pm 0.028}$ | $\mathbf{0.479 \pm 0.045}$ | $\mathbf{0.621 \pm 0.009}$ |
| CLARI-M | 400 | 200 | $0.879 \pm 0.023$ | $0.506 \pm 0.028$ | $0.646 \pm 0.007$ |
| CLARI-L | 400 | 200 | $\mathbf{0.919 \pm 0.018}$ | $0.596 \pm 0.039$ | $0.669 \pm 0.007$ |
| CLARI-H | 400 | 200 | $0.891 \pm 0.020$ | $\mathbf{0.678 \pm 0.032}$ | $\mathbf{0.771 \pm 0.007}$ |
| CLARI-M | 1000 | 1000 | **0.960** | 0.620 | 0.754 |
| CLARI-L | 1000 | 1000 | 0.940 | 0.760 | 0.763 |
| CLARI-H | 1000 | 1000 | **0.960** | **0.780** | **0.856** |

| Method | $n_s$ | $k$ | CSP5 (6) | CSP6 (5) | CSP7 (8) |
|---|---|---|---|---|---|
| $\text{DFT}_{\text{avg}}$ | – | – | 0.544 | 0.496 | 0.421 |
| OXtal | 30 | 30 | 0.167 | 0.200 | 0.125 |
| CLARI-M | 30 | 30 | $0.554 \pm 0.140$ | $0.311 \pm 0.134$ | $0.210 \pm 0.089$ |
| CLARI-M | 150 | 30 | $0.616 \pm 0.134$ | $0.374 \pm 0.153$ | $0.225 \pm 0.080$ |
| CLARI-L | 30 | 30 | $0.681 \pm 0.140$ | $0.355 \pm 0.148$ | $0.245 \pm 0.096$ |
| CLARI-L | 150 | 30 | $\mathbf{0.789 \pm 0.142}$ | $0.480 \pm 0.155$ | $0.263 \pm 0.091$ |
| CLARI-H | 30 | 30 | $0.705 \pm 0.120$ | $0.445 \pm 0.149$ | $0.280 \pm 0.119$ |
| CLARI-H | 150 | 30 | $0.785 \pm 0.113$ | $\mathbf{0.612 \pm 0.152}$ | $\mathbf{0.393 \pm 0.137}$ |
| CLARI-M | 400 | 200 | $0.863 \pm 0.099$ | $0.657 \pm 0.139$ | $0.384 \pm 0.066$ |
| CLARI-L | 400 | 200 | $\mathbf{0.975 \pm 0.061}$ | $0.729 \pm 0.104$ | $0.566 \pm 0.128$ |
| CLARI-H | 400 | 200 | $0.965 \pm 0.069$ | $\mathbf{0.766 \pm 0.077}$ | $\mathbf{0.822 \pm 0.126}$ |
| CLARI-M | 1000 | 1000 | **1.000** | **0.800** | 0.500 |
| CLARI-L | 1000 | 1000 | **1.000** | **0.800** | 0.875 |
| CLARI-H | 1000 | 1000 | **1.000** | **0.800** | **1.000** |

Table 8: Extended OXtal metrics on the OXtal benchmark subsets. When $k < n_s$, our methods select the top-$k$ by UMA energy.

| Subset | Method | $n_s$ | $k$ | $\text{Col}_S \downarrow$ | $\text{Pac}_S \uparrow$ | $\text{Pac}_C \uparrow$ | $\text{Rec}_S \uparrow$ | $\text{Rec}_C \uparrow$ |
|---|---|---|---|---|---|---|---|---|
| Rigid | OXtal | 30 | 30 | 0.011 | 0.873 | 1.000 | 0.737 | 0.960 |
| | CLARI-M | 30 | 30 | 0.000 | 0.276 | 0.787 | 0.921 | 0.992 |
| | CLARI-L | 30 | 30 | 0.000 | 0.292 | 0.819 | 0.923 | 1.000 |
| | CLARI-H | 30 | 30 | 0.000 | 0.345 | 0.822 | 0.929 | 0.999 |
| | CLARI-M | 150 | 30 | 0.000 | 0.303 | 0.802 | 0.918 | 0.986 |
| | CLARI-L | 150 | 30 | 0.000 | 0.323 | 0.852 | 0.924 | 1.000 |
| | CLARI-H | 150 | 30 | 0.000 | 0.387 | 0.833 | 0.930 | 0.997 |
| Flexible | OXtal | 30 | 30 | 0.097 | 0.291 | 0.900 | 0.048 | 0.400 |
| | CLARI-M | 30 | 30 | 0.000 | 0.034 | 0.347 | 0.273 | 0.828 |
| | CLARI-L | 30 | 30 | 0.000 | 0.042 | 0.394 | 0.266 | 0.805 |
| | CLARI-H | 30 | 30 | 0.000 | 0.064 | 0.458 | 0.339 | 0.869 |
| | CLARI-M | 150 | 30 | 0.000 | 0.041 | 0.369 | 0.272 | 0.826 |
| | CLARI-L | 150 | 30 | 0.000 | 0.056 | 0.441 | 0.278 | 0.804 |
| | CLARI-H | 150 | 30 | 0.000 | 0.094 | 0.553 | 0.377 | 0.884 |
| CSP5 | OXtal | 30 | 30 | 0.006 | 0.667 | 0.833 | 0.572 | 0.833 |
| | CLARI-M | 30 | 30 | 0.000 | 0.087 | 0.639 | 0.826 | 0.891 |
| | CLARI-L | 30 | 30 | 0.000 | 0.103 | 0.749 | 0.816 | 0.938 |
| | CLARI-H | 30 | 30 | 0.000 | 0.252 | 0.737 | 0.822 | 0.913 |
| | CLARI-M | 150 | 30 | 0.000 | 0.097 | 0.671 | 0.828 | 0.903 |
| | CLARI-L | 150 | 30 | 0.000 | 0.114 | 0.828 | 0.824 | 0.940 |
| | CLARI-H | 150 | 30 | 0.000 | 0.286 | 0.807 | 0.830 | 0.883 |
| CSP6 | OXtal | 30 | 30 | 0.013 | 0.660 | 1.000 | 0.160 | 0.600 |
| | CLARI-M | 30 | 30 | 0.000 | 0.200 | 0.402 | 0.544 | 0.868 |
| | CLARI-L | 30 | 30 | 0.000 | 0.200 | 0.445 | 0.556 | 0.908 |
| | CLARI-H | 30 | 30 | 0.000 | 0.211 | 0.521 | 0.572 | 0.917 |
| | CLARI-M | 150 | 30 | 0.000 | 0.208 | 0.533 | 0.565 | 0.911 |
| | CLARI-L | 150 | 30 | 0.000 | 0.210 | 0.512 | 0.569 | 0.868 |
| | CLARI-H | 150 | 30 | 0.000 | 0.227 | 0.647 | 0.588 | 0.943 |
| CSP7 | OXtal | 30 | 30 | 0.021 | 0.483 | 0.875 | 0.129 | 0.375 |
| | CLARI-M | 30 | 30 | 0.000 | 0.066 | 0.324 | 0.410 | 0.706 |
| | CLARI-L | 30 | 30 | 0.000 | 0.071 | 0.366 | 0.394 | 0.728 |
| | CLARI-H | 30 | 30 | 0.000 | 0.080 | 0.419 | 0.400 | 0.738 |
| | CLARI-M | 150 | 30 | 0.000 | 0.071 | 0.307 | 0.419 | 0.712 |
| | CLARI-L | 150 | 30 | 0.000 | 0.075 | 0.332 | 0.392 | 0.733 |
| | CLARI-H | 150 | 30 | 0.000 | 0.111 | 0.521 | 0.436 | 0.780 |

### D.2 Polymorph coverage

To evaluate whether CLARI recovers multiple polymorphs, we collect the 48 refcode families from the OXtal test set that have multiple members. For each family, we draw $n_s = 150$ samples from CLARI-H and match them against all members. The coverage of a family is the fraction of members that are matched (in the Sol sense) by at least one sample, which we average over families uniformly or size-weighted. CLARI-H recovers 61–75% of the observed polymorphs (Table 9) using only 150 samples per family without energy ranking, whereas template generation approaches routinely generate thousands of samples before refinement and down-selection [104]. Coverage is stronger on the rigid families than on the flexible and blind-test families, showing room for future work. Samples are also diverse: across the 48 families, the family-balanced mean within-family EMD PDD is 0.208 for experimental polymorphs and 0.412 for 1000 CLARI-H samples per target. Table 10 reports the number of clusters among these samples at different EMD PDD thresholds.

Table 9: Polymorph coverage of CLARI-H on the 48 OXtal test refcode families with multiple members, using $n_s = 150$ samples per family without energy ranking. Coverage is averaged over families uniformly or size-weighted.

| | | | Family size | | Coverage | |
|---|---|---|---|---|---|---|
| Group | Families | Polymorphs | Median | [Min, Max] | Uniform | Weighted |
| Rigid | 30 | 379 | 4 | [2, 74] | 0.757 | 0.871 |
| Flexible | 10 | 181 | 12.5 | [3, 53] | 0.376 | 0.564 |
| CSP5 | 1 | 13 | 13 | [13, 13] | 0.231 | 0.231 |
| CSP6 | 2 | 7 | 3.5 | [2, 5] | 0.300 | 0.429 |
| CSP7 | 5 | 11 | 2 | [2, 3] | 0.400 | 0.364 |
| All | 48 | 591 | 4.5 | [2, 74] | 0.611 | 0.748 |

Table 10: Median number of clusters among 1000 CLARI-H samples per OXtal test target at different EMD PDD thresholds. Cluster counts are threshold-dependent and should not be interpreted as numbers of physical polymorphs.

| EMD PDD threshold | 0.10 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.50 |
|---|---|---|---|---|---|---|---|
| Median clusters | 1000 | 978 | 851 | 525 | 157 | 38 | 2 |

### D.3 Supplementary figures

In the following pages, we provide additional supporting figures.

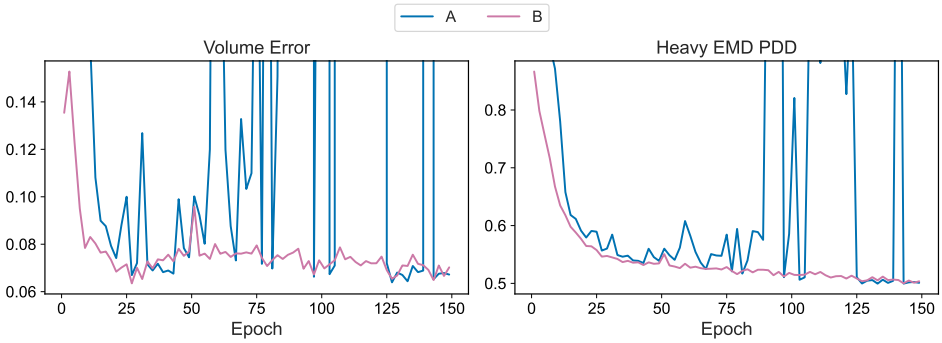



Figure 7: Reconstruction metric curves on the validation set for models **A** and **B** (Table 2). Results were pulled from preliminary training logs, where we only evaluated on a subset of 512 validation crystals, drew 3 samples per crystal, and did a sample-wise mean (not min.) aggregation.

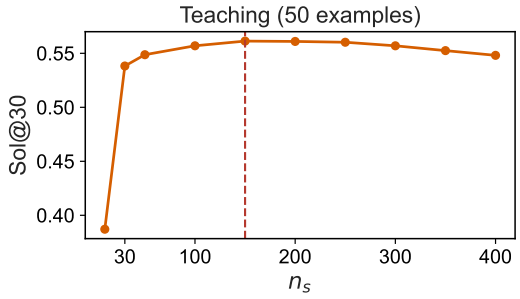
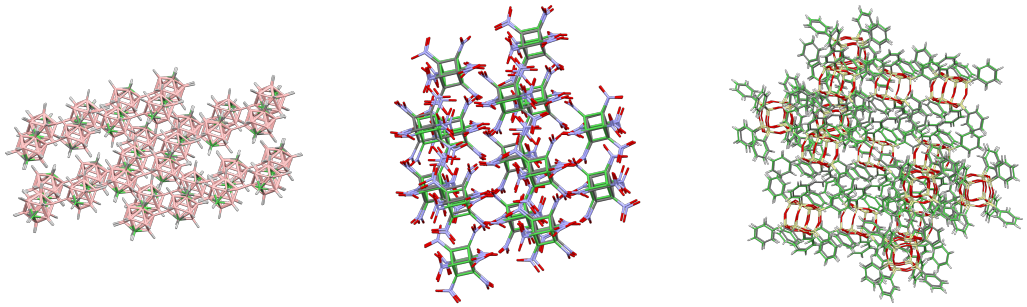
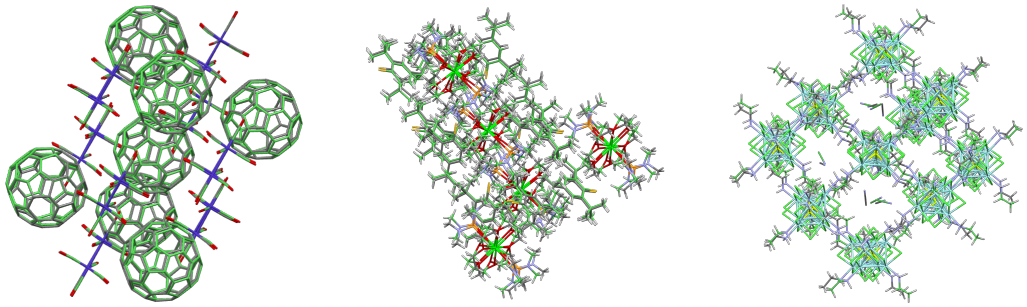


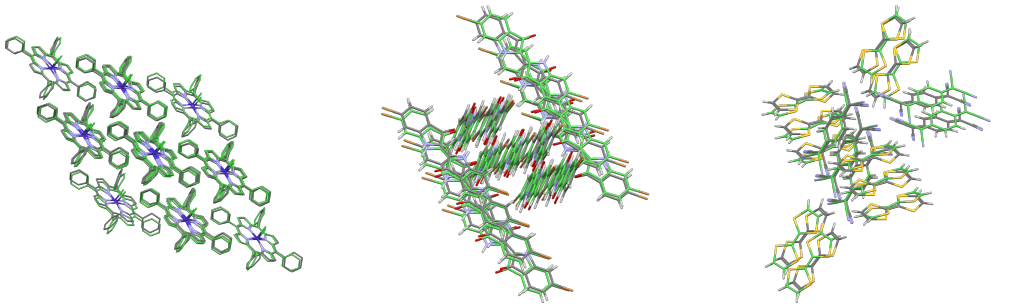

Figure 8: Sol@30 as a function of $n_s$ on 50 random examples drawn from the CSD Teaching Subset. We find that $n_s = 150$ achieves the best solve rate for Sol@30. We sample 1000 times per example and report the mean over 5000 bootstrap resamples.



(a) BCABOR10, 0.117 Å (b) CUGDIR, 0.356 Å (c) DORRAF, 0.252 Å



(d) YUGWUT, 0.226 Å (e) JIYDEA, 0.575 Å (f) SIMLIJ, 1.082 Å



(g) CTPOCO, 0.445 Å (h) DBRING01, 0.442 Å (i) TTFTCQ, 0.739 Å

Figure 9: Predicted crystal structures from the CSD Teaching Subset, all achieving full 15/15 molecule matches in COMPACK with the reported $RMSD_{15}$ values (Å).

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims of the paper are that modelling only the unit cell and replacing triangle layers with pair-bias improves generative modelling of organic crystal structure. The claims are substantiated by results showing a higher solve rate at a lower computational cost.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations, such as being agnostic to chirality, as well as the limitations of the Sol metric, are discussed in Section 5.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in

developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not include theoretical results.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The architecture, training losses, training recipe, hyperparameters, data split, and evaluation protocol are documented throughout the main text and appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We release code at https://github.com/the-matter-lab/clari and describe all details of experiments in the Appendix. However, we train on the Cambridge Structural Database, which is commercially licensed from the CCDC.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section A describes the data processing pipeline and data splits in detail. Table 5 in Section B.6 lists the hyperparameters.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report bootstrapped means of solve rate via resampling, as well as standard errors in Table 7.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Table 5 in Section B.6 describes the compute resources and timing used for the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

Justification: The work conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: A discussion of the positive and dual-use risks of the work is given in Section 5.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: Despite the dual-use risk of CLARI for developing organic explosives, it does not have a higher risk of misuse than accessing the original database.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The Cambridge Structural Database is licensed commercially and cited. Other models (OXtal, UMA) and methods (COMPACK, PDD) are cited at first use, and are used within their public terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code introduced by the paper is documented in the repository. Training on CSD data is within the terms of the license.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.