OXTAL: AN ALL-ATOM DIFFUSION MODEL FOR ORGANIC CRYSTAL STRUCTURE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurately predicting experimentally-realizable 3D molecular crystal structures from their 2D chemical graphs is a long-standing open challenge in computational chemistry called *crystal structure prediction* (CSP). Efficiently solving this problem has implications ranging from pharmaceuticals to organic semiconductors, as crystal packing directly governs the physical and chemical properties of organic solids. In this paper, we introduce OXTAL, a large-scale 100M parameter all-atom diffusion model that directly learns the conditional joint distribution over intramolecular conformations and periodic packing. To efficiently scale OXTAL, we abandon explicit equivariant architectures imposing inductive bias arising from crystal symmetries in favor of data augmentation strategies. We further propose a novel crystallization-inspired lattice-free training scheme, STOICHIOMETRIC STOCHASTIC SHELL SAMPLING (S^4), that efficiently captures long-range interactions while sidestepping explicit lattice parametrization—thus enabling more scalable architectural choices at all-atom resolution. By leveraging a massive dataset of 600K experimentally validated crystal structures (including rigid and flexible molecules, co-crystals, and solvates), OXTAL achieves orders-of-magnitude improvements over prior *ab-initio* machine learning CSP methods, while remaining orders of magnitude cheaper than traditional quantum-chemical approaches. Specifically, OXTAL recovers experimental structures with conformer $\mathrm{RMSD}_1 < 0.5$ Å and attains over 80% lattice-match success, demonstrating its ability to model both thermodynamic and kinetic regularities of molecular crystallization.

1 INTRODUCTION

A landmark open challenge in computational chemistry is identifying a molecule's 3D crystal structure given knowledge of its chemical composition (Bardwell et al., 2011; Reilly et al., 2016; Hunnisett et al., 2024). In particular, given only a molecule's 2D chemical graph, *ab initio* molecular crystal structure prediction (CSP) seeks to estimate the distribution of experimentally realizable crystal packings in an accurate and scalable manner. This 3D arrangement of molecules within a periodic lattice dictates the macroscopic behavior of organic solids. For instance, in pharmaceuticals, crystal packing governs solubility, bioavailability, and the long-term stability of active ingredients (Schultheiss & Newman, 2009; Chen et al., 2011); in material science, intermolecular geometry dictates charge transport, porosity, and optical response—enabling applications across electronics, photonics, sensing, and energy storage (Zhang et al., 2018; Wang et al., 2019).

The complexity of CSP is underscored by the nature of crystal formation, wherein experimentally realized structures often occupy local minima of a highly non-smooth and well-separated Gibbs free energy landscape (Figure 3(a)). This thermodynamic energy landscape is determined by the competition between the *intramolecular* interactions that set the molecule's own (flexible) conformation and the long-range and weak *intermolecular* forces that dictate how molecules pack together periodically (Chernov, 2012). As a result, classical CSP approaches combine a search procedure (e.g., enumeration or evolutionary algorithms) with oracle access to energy/ranking models, such as force fields or quantum-chemical density functional theory (DFT) (Engel & Dreizler, 2011; Hunnisett et al., 2024). However, classical CSP approaches often fail to capture realistic *kinetic* conditions that lead to the *distribution* of experimentally sampled energy minima. Consequently, these methods require the generation and optimization of ~1,000 to 100,000 structures per molecule—the majority of which struggle to go beyond unfavourable local energy minima despite extensive computation.

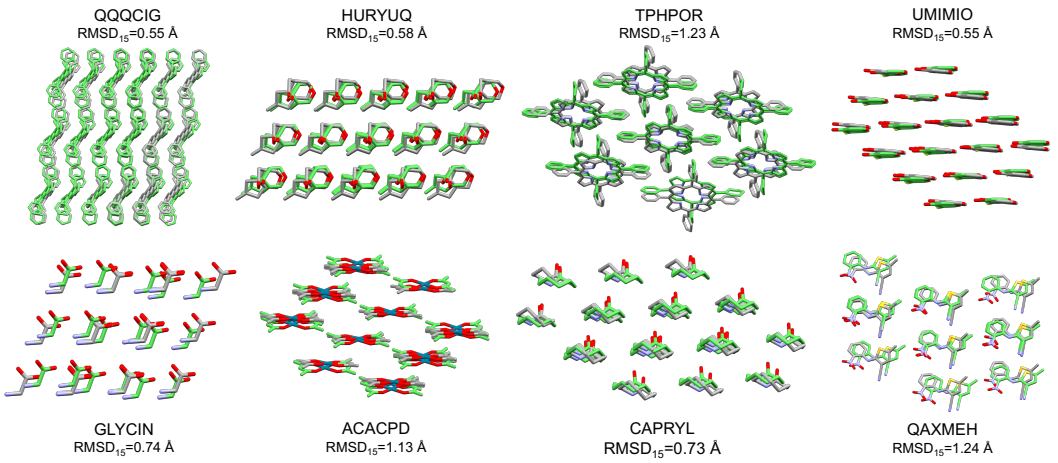


Figure 1: Molecular crystal structures generated by OXTAL (color) compared to ground truth (grey).

Recently, *generative* approaches to modelling atomic systems–e.g., AlphaFold3 (Abramson et al., 2024) for biomolecules and MatterGen (Zeni et al., 2025) for inorganic materials–have demonstrated their ability to capture intricate 3D atomistic interactions directly from data. Molecular CSP generalizes both of these, as proteins and inorganic crystals have a smaller set of interactions; proteins pack into lattices under strong intramolecular constraints mostly governed by their backbones (Branden & Tooze, 2012), and inorganic crystals possess strong covalent or ionic bonds across a smaller number of atoms (Figure 2a). Protein generative models also heavily rely on biological priors based on evolutionary information captured by multiple-sequence alignment (Jumper et al., 2021). In contrast, small-molecule crystals span chemically diverse scaffolds, exhibit rich conformational flexibility, and often contain many molecular copies within a unit cell (Figure 2b). Accurately capturing these interactions requires a large and diverse training set, highly expressive yet efficient to sample models, and training schemes that consider periodic interactions in a crystal lattice without impeding scalability for large model training.

Main contributions. In this paper, we introduce OXTAL, an all-atom diffusion transformer model for molecular CSP. Conditioned solely on the 2D molecular graph, OXTAL learns to sample experimentally realistic crystal structures with accurate descriptions of molecular conformers as well as their periodic packing. To enable large-scale modelling, we train OXTAL on over 600k experimental molecular crystal structures spanning rigid and flexible molecules, co-crystals, and solvates. OXTAL builds on recent advances in all-atom generative modelling (Abramson et al., 2024), discarding symmetry representations of lattice vectors and associated crystal symmetries in favor of training directly on Cartesian coordinates and employing $\mathrm{SE}(3)$ data augmentation. We further introduce STOICHIOMETRIC STOCHASTIC SHELL SAMPLING (S^4), a novel lattice-free training scheme that retains the rich information of long-range interactions. We summarize our key contributions as follows:

- We present OXTAL, the first large-scale all-atom diffusion model for molecular CSP, that samples molecular crystal packing directly from 2D molecular graphs (§3).
- We introduce a crystallization-inspired training scheme for periodic structures, S^4, which removes explicit lattice parametrization and enables more scalable training (§3.1).
- Empirically, OXTAL significantly outperforms existing ML-based *ab initio* CSP methods, achieving $\mathrm{RMSD}_1 < 0.5$Å and lattice match rates above 80% within 30 samples (§4.1), while being several orders of magnitude cheaper than traditional quantum chemical methods in DFT (§4.2).
- We provide additional chemical analysis highlighting OXTAL's ability to capture diverse intra-/intermolecular interactions, including crystal polymorphs, and generalize to complex co-crystal and biomolecular interactions (§4.3).

2 BACKGROUND AND PRELIMINARIES

2.1 CRYSTAL REPRESENTATIONS

Formally, a periodic crystal structure $\mathcal{C}$ is defined by a pair $(L, \mathcal{B})$. The first component $L \in \mathbb{R}^{3\times3}$ defines the lattice vectors forming a parallelepiped known as the unit cell. The second component $\mathcal{B} = \{(z_i, u_i)\}_{i=1}^{N}$ is the basis, which consists of the N atoms within this unit cell. Each atom

is described by its species $\{z_i\}_{i=1}^N$ and its fractional coordinate $\{u_i \in [0,1)^3\}_{i=1}^N$ relative to the lattice vectors, or equivalently, its Cartesian coordinate Lu_i. For molecular crystals, $\mathcal{B}$ naturally decomposes into Z molecules. The connectivity of each molecule m_k is given by a graph $\{g_k = (V_k, E_k)\}_{k=1}^Z$ (vertices/atoms V_k and edges E_k; species labels z_i are carried by the vertices). We next recall the symmetries present in the periodic crystal structure $\mathcal{C}$.

Symmetries. Given $\mathcal{C} = (L, \mathcal{B})$ where $(u_i \in \mathbb{T}^3 := \mathbb{R}^3/\mathbb{Z}^3)$, the Cartesian positions of atoms are,

$$X(L, \mathcal{B}) = \{\, L(n + u_i) : \; n \in \mathbb{Z}^3, \; i = 1, \ldots, N \,\}. \tag{1}$$

Furthermore, two descriptions $(L, \mathcal{B})$ and $(L', \mathcal{B}')$ encode the same structure if and only if there is a group action g in 3D, $g := (R, t) \in \mathrm{SE}(3)$ such that, $X(L', \mathcal{B}') = g \circ X(L, \mathcal{B})$.

A periodic crystal admits an *asymmetric unit* $\mathcal{A}$, the minimal subset that recovers the entire unit cell by applying symmetry transformations of the crystal's space group. Conversely, a supercell can be obtained by an integer matrix $U \in \mathbb{Z}^{3\times3}$ with $m := |\det U| \geq 1$, yielding $\mathcal{C}^{(U)} = (LU, \mathcal{B}^{(U)})$, the same infinite crystal in a differing tiling. For example, $U = mI_3$ yields a cubic $m \times m \times m$ supercell. A formal summary of crystal symmetries is outlined below.

Crystal representation invariances

(S1) *Global translation.* For $t \in \mathbb{T}^3$, $(L, \{(z_i, u_i)\}) \mapsto (L, \{(z_i, u_i + t)\})$.

(S2) *Global rotation.* For $R \in \mathrm{SO}(3)$, $(L, \{(z_i, u_i)\}) \mapsto (RL, \{(z_i, u_i)\})$.

(S3) *Permutation (reindexing).* For $\zeta \in S_N$,$(L, \{(z_i, u_i)\}) \mapsto (L, \{(z_{\zeta(i)}, x_{\zeta(i)})\})$.

(S4) *Unit-cell change.* Let $U \in \mathbb{Z}^{3\times3}$ with $m := |\det U| \geq 1$.

- *Unimodular basis change* ($m = 1, U \in \mathrm{GL}(3, \mathbb{Z})$):
$$(L, \{(z_i, u_i)\}) \mapsto (LU, \{(z_i, U^{-1}u_i)\}), \quad (LU)(n + U^{-1}u_i) = L(Un + u_i).$$
- *Supercell expansion* ($m > 1$): let $\mathcal{R}(U) \subset \mathbb{T}^3$ be a fixed set of coset representatives for $\mathbb{Z}^3/U\mathbb{Z}^3$ (so $|\mathcal{R}(U)| = m$). Then
$$(L, \{(z_i, u_i)\}) \mapsto \Big(LU, \; \{(z_i, U^{-1}(u_i + r)) : \; i = 1, \ldots, N, \; r \in \mathcal{R}(U)\}\Big),$$
since $(LU)(n + U^{-1}(u_i + r)) = L(Un + u_i + r)$.

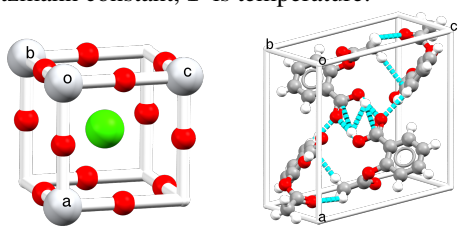
Molecular crystallization. Let $g = \{g_k = (V_k, E_k)\}_{k=1}^Z$ denote the set of molecular graph(s) and $\mathcal{X}(g)$ the set of periodic, all-atom crystal structures compatible with g. Due to the invariances, the physically distinct configurations form a quotient space $\mathcal{M}(g) = \mathcal{X}(g)/\sim$. *Ab initio* CSP can then be posed as conditional probabilistic inference over equivalence classes $[\,\mathcal{C}\,] \equiv [\,(L, \mathcal{B})\,] \in \mathcal{M}(g)$. An approximation of the experimentally realized distribution under crystallization conditions $\aleph$ is:

$$p_\aleph\big([\,\mathcal{C}\,] \mid g\big) \;\propto\; \kappa_\aleph\big([\,\mathcal{C}\,] \mid g\big)\, \exp\big(-\beta\, \Delta G([\,\mathcal{C}\,])\big), \qquad \beta = \tfrac{1}{k_B T} \tag{2}$$

where ΔG represents the Gibbs free energy (thermodynamics), $\kappa_\aleph$ summarizes kinetic accessibility (nucleation and growth pathways), and k_B is the Boltzmann constant, T is temperature.

In practice, learning and sampling must respect the symmetry of $\mathcal{M}(g)$, handle strong coupling between intramolecular conformation and intermolecular packing (especially in flexible molecules), navigate rugged kinetic and energy landscapes arising from large unit cells (often > 100 atoms) and weak and long-range interactions, and marginalize over unknown Z.

These challenges are distinct from most inorganic crystals, which are often characterized by strong covalent or ionic bonds, contain < 30 atoms per unit cell, and lack molecular conformers (Figure 2).



(a) $CaTiO_3$ (inorganic) (b) Asprin (organic)

Figure 2: Molecular crystals consist of distinct molecules held together via long-range, weak interactions. They typically contain many atoms per unit cell and unknown molecule copies Z.

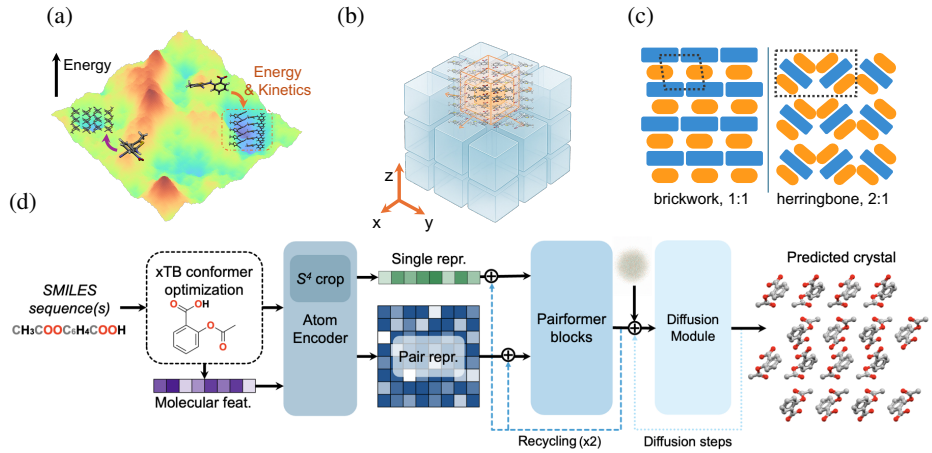


Figure 3: (a) Schematic of a rugged crystallization Gibbs free energy landscape with many local minima. Kinetic conditions often dictate which experimental minimum is formed. (b) Molecular crystallization, showing *nucleation* and *growth* in successive layers, which is the inspiration for S^4. (c) Common packing motifs exemplified in co-crystal polymorphs with 1:1 and 2:1 stoichiometric ratio. (d) Overview of OXTAL architecture.

2.2 CONTINUOUS-TIME DIFFUSION MODELS

A diffusion model solves the generative modelling problem by being the solution to a (Itô) stochastic differential equation (SDE) (Øksendal, 2003),

$$\mathrm{d}\mathbf{X}_t = f_t(\mathbf{X}_t)\,\mathrm{d}t + \sigma_t \mathrm{d}\mathbf{W}_t, \quad \mathbf{X}_0 \sim p_0. \tag{3}$$

In Equation (3), we use boldface to denote random variables and normal script to denote functions or samples. The function $f_t : \mathbb{R}^d \to \mathbb{R}^d$ corresponds to the average direction of evolution, while $\sigma_t : \mathbb{R} \to \mathbb{R}$ is the diffusion coefficient for the Wiener process $\mathbf{W}_t$. By convention, this is termed the forward noising process that starts from time $t = 0$ and progressively corrupts datae until the terminal time $t = 1$ where we reach a structureless prior $p_1(x_1) := \mathcal{N}(0, I)$.

Under mild regularity assumptions, forward SDEs of the form of Equation (3) admit a time reversal, which itself is another SDE in the reverse direction $t = 1$ to $t = 0$ and transmutes a prior sample $x_1 \sim p_1$ to a sample from the target distribution $x_0 \sim p_0$ (Anderson, 1982),

$$\mathrm{d}\mathbf{X}_t = \left(f_t(\mathbf{X}_t)\,\mathrm{d}t - \sigma_t^2 \nabla_x \log p_t(\mathbf{X}_t)\right)\,\mathrm{d}t + \sigma_t \mathrm{d}\overline{\mathbf{W}}_t, \quad \mathbf{X}_1 \sim p_1. \tag{4}$$

Here $\overline{\mathbf{W}}_t$ is another standard Wiener process, and the *reverse-time* SDE shares the same time marginal density p_t as the forward SDE. Critically, the forward and reverse SDEs are linked via the Stein score $\nabla_x \log p_t(x_t)$, which is the key quantity of interest in the design of diffusion models. More precisely, diffusion models estimate the score function by forming an ℓ_2-regression objective using a denoising network $D_\theta(x_t, t)$. For example, for the variance exploding (VE) SDE family of forward processes: $p_t = p_0 * \mathcal{N}(0, \sigma_t^2)$, this corresponds to learning the set of optimal denoisers, i.e., the set of conditional expectations, across time $D(x_t, t) = \mathbb{E}\left[\mathbf{X}_0 | \mathbf{X}_t = x_t\right], \forall t \in [0, 1]$. Owing to the nature of Gaussian convolution, we can convert this to a simple *simulation-free* training *conditional* objective for any Bregman divergence with convex F, $\mathbb{B}_\mathrm{F}$ (Holderrieth et al., 2025, Prop. 2)

$$\mathcal{L}_c(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,1), x_0\sim p(x_0), x_t\sim p_t(x_t|x_0)}\left[\lambda(t)\mathbb{B}_\mathrm{F}(x_0, D_\theta(x_t, t))\right],$$

where $\lambda(t) : [0, 1] \to \mathbb{R}_+$ is any weighting function. For a sufficiently expressive family of denoisers $\mathcal{H} = \{D_\theta : \theta \in \Theta\}$, the minimizer to Section 2.2 is $D^*_\theta(x_t, t) = D(x_t, t) = \mathbb{E}\left[\mathbf{X}_0 | \mathbf{X}_t = x_t\right]$. Given a trained denoising model, diffusion models generate samples at inference by simulating the reverse-time dynamics of §4 with the learned score s_θ computed as $s_\theta \leq (D_\theta(x_t, t) - x_t)/\sigma_t^2$.

3 OXTAL

We now describe OXTAL, our all-atom diffusion model for molecular CSP. As outlined in §2.1, existing inorganic CSP models (Miller et al., 2024; Gasteiger et al., 2021) that rely on equivariant architectures and explicit unit-cell parametrizations face scalability challenges for large molecular crystals with unknown multiplicity Z. We thus introduce S^4 training, which exposes the model to long-range periodic cues without ever parametrizing a lattice (§3.1). Second, we implement a high-

capacity, non-equivariant Transformer with data augmentation and strong molecular embeddings to capture symmetries (§3.2). Together, these choices decouple what to generate (conformations and packing) from how the crystal is represented during training (unit cell, Z, etc.).

3.1 Stoichiometric Stochastic Shell Sampling (S^4)

Crystallization is a local-to-global process: once molecules approach contact distances, weak but specific interactions induce recurring motifs that propagate periodically. Learning to denoise such local consistent neighborhoods should therefore recover larger periodicity at inference time. Training on such subsampled blocks reduces token size, provides natural augmentation, and mirrors the partial observability of nucleation and growth (Figure 3(b)).

We formalize this idea in S^4. Let $\mathcal{C}^{(U)} = (LU, \mathcal{B}^{(U)})$ be a supercell with molecules $m \in \mathcal{M}$, where $X(m) \subset \mathbb{R}^3$ denotes m's non-hydrogen Cartesian coordinates. We next define the *minimum-image* intermolecular distance between two molecules, $d_{\min}(m, m') = \min_{x \in X(m),\, x' \in X(m')} \|x - x'\|_2$.

Given a fixed contact radius r_{cut}, S^4 builds concentric shells $\mathcal{S}$ around a uniformly-sampled central molecule $m_c \sim \text{Uniform}(\mathcal{A})$ based on the molecular contact graph induced by $d_{\min}(\cdot,\cdot) \le r_{\text{cut}}$:

$$\mathcal{S}_k(m_c) = \{m_i \in \mathcal{M} : (k-1)\epsilon \le d_{min}(m_c, m_i) < (k)\epsilon\}, \quad k = 1, 2, \ldots \tag{5}$$

We sample the number of shells $K \sim \text{Uniform}(1, k_{max})$, and define the block of molecules $V_K = \cup_{j=0}^{K} S_j$. We cap block size by a token budget $T_{\max}$ (atoms). If $|V_K| \le T_{\max}$ we accept V_K; otherwise we choose the smallest $K^\star$ with $|V_{K^\star+1}| > T_{\max}$ and subsample the frontier $S_{K^\star}$ to meet the budget while preserving the crystal's molecular stoichiometry. For example, if molecule type i appears N_i^{ASU} times in $\mathcal{A}$ and N_i times in $S_{K^\star}$, we sample molecules of type i in $S_{K^\star}$ with weight $\omega_i \propto (N_i^{\text{ASU}}/N_i)$. The resulting set is our training crop, denoted $\mathbf{A}_{\text{crop}}$.

Compared to centroid- or kNN-based heuristics, this "shell cropping" better respects local interaction networks by respecting anisotropic packing motifs and interactions beyond the strongest ones (Figure 9), while mitigating truncation biases common in large-graph learning (Zeng et al., 2019). Appendix D.1 shows that training with S^4 outperforms kNN and centroid cropping methods. Appendix F.1 show that training with S^4 can generalize to long-range periodicity beyond token sizes in training. Hyperparameter ablations of r_{cut} can be found in Appendix D.2. Finally, we bound the error due to cropping. In the next proposition, we show that the error in the loss due to cropping with S^4 decreases with the cube root of the number of tokens.

Proposition 1. *Let $\partial\mathbf{A}_{crop} = \{\{u, v\} \in E : u \in \mathbf{A}_{crop}, v \notin \mathbf{A}_{crop}\}$ represent the boundary of $\mathbf{A}_{crop}$. Denote the number of atoms in a volume C as $T(C)$. Let $L_\partial(\mathbf{A}_{crop}) = \sum_{\{u,v\} \in \partial\mathbf{A}_{crop}} \mathcal{L}(u, v)$ be the boundary loss. Assuming $\exists r^0$ s.t. $L(u, v) = 0, \forall u, v$ s.t. $\|u - v\| > r^0$, i.e. $\mathcal{L}$ is local, and there exist $0 < a \le b < \infty$ s.t. $a|S| \le T(S) \le b|S|$ for any $S \subseteq V$. Then,*

$$\frac{L_\partial(\mathbf{A}_{crop})}{T(\mathbf{A}_{crop})} = O((1+\epsilon)T(\mathbf{A}_{crop})^{-1/3}) \tag{6}$$

3.2 Model Architecture

OXTAL is comprised of: (1) a **Molecular Encoder** that embeds both physical and structural information; (2) a **Pairformer Trunk** that propagates information across all atoms in the crop; (3) a **Diffusion Module** that takes in the single and pairwise representations and outputs a generated crystal structure. The overall architecture of OXTAL is depicted in Figure 3(d).

Molecular encoder. Given an input SMILES sequence s, we generate a 3D conformer with RDKit ETKDG followed by relaxation by the semi-empirical quantum chemical method GFN2-xTB (Pracht et al., 2020). The atomic number, positions, formal charges, Mulliken partial charges, and bond information are used as *feature embeddings* for the model. OXTAL is not particularly sensitive to the feature conditioning conformer coordinates (Appendix E). Finally, we resolve ambiguity of identical molecular copies via relative position encoding on entity identifiers (Abramson et al., 2024).

Pairformer trunk. We adapt existing state-of-the-art architectures for protein folding to generate single and pair representations for each molecular crystal (Abramson et al., 2024). Instead of tokenizing protein residues, we simplify the tokenization such that each token directly represents a single atom a_i in the molecule. We then apply the Pairformer Stack from AlphaFold 3, which leverages triangular self-attention to update the single and pair representations. Unlike AlphaFold

2, which relied on the equivariant Evoformer (Jumper et al., 2021) architecture, this simpler Pairformer module is not explicitly equivariant, allowing for training on larger sequences.

Diffusion module. The design of our diffusion module follows that of AlphaFold3 (Abramson et al., 2024), consisting of an atom attention encoder which combines token information given by the pairformer with an encoded representation of x_t, followed by a large 70M parameter diffusion transformer (Peebles & Xie, 2023), before a final atom attention decoder which predicts the denoised atomic positions. We broadly follow Karras et al. (2022) for pre-conditioning of model inputs. In Appendix D.3, we compare the non-equivariant transformer to the SE(3)-equivariant EquiformerV2 (Liao et al., 2024). Under a matched memory budget, the substantially higher memory consumption of EquiformerV2 restricts the model size, leading to markedly degraded performance. Diffusion module size ablation is further provided in Appendix D.4.

3.3 Training

OXTAL is trained using a procedure similar to those successfully employed for protein structure prediction (Abramson et al., 2024). The training objective is a composite loss designed to capture both the global structure and the accuracy of the local chemical environment. This loss is comprised of two main components: (1) a mean squared error loss $\mathcal{L}_{\text{mse}}$, (2) a smooth local difference distance test $\mathcal{L}_{\text{sLDDT}}$ as defined in Abramson et al. (2024). Both losses compare the predicted structure $\hat{x}_0 := D_\theta(x_t, t)$ and an aligned ground truth structure $x_0^{\text{align}} = \texttt{align}(x_0, \hat{x}_0)$, and the latter emphasizes on the pairwise interactions within the crop via a surrogate of the interatomic distances. To round off our training, we also include a distogram loss on a separate head branching from the trunk $\mathcal{L}_{\text{dist}}(\hat{d}, d)$ to ensure the trunk output contains binned pairwise distance information. Additional details for computing component losses are provided in Appendix B.2.3. The final loss is then a weighted sum of these components:

$$\mathcal{L}(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,1), x_t\sim p_t\left(x_t|x_0^{\text{align}}\right)}\left[\mathcal{L}_{\text{mse}}(\hat{x}_0, x_0^{\text{align}}) + \mathcal{L}_{\text{sLDDT}}(\hat{x}_0, x_0^{\text{align}})\right] + \lambda_{\text{dist}}\mathcal{L}_{\text{dist}}(\hat{d}, d). \quad (7)$$

We next curate a training dataset from the Cambridge Structural Database (CSD) that contains $\sim$ 600k crystals. Specific details regarding model training and configuration are outlined in §B.

4 Experiments

We evaluate OXTAL on several different datasets for molecular CSP, comparing against a range of ML-based (§4.1) and DFT energy-based methods (§4.2). These results are complemented with a broader chemical survey (§4.3). See §C for exact specifications.

Baselines. We compare against ML *ab initio* methods and energy-based methods. For ML methods, as most models for inorganic CSP are incompatible, we evaluate the pre-trained AssembleFlow, a molecular CSP method that infers crystal packing from rigid-body molecules (Guo et al., 2025) and A-Transformer, an all-atom transformer flow matching model (§C.3.1). Additional results for zero-shot inference from AlphaFold3 are presented in §C.3.3. For energy-based methods, we compare against computational chemistry baselines that submitted to CCDC's 5th, 6th, and 7th CSP blind tests (Bardwell et al., 2011; Reilly et al., 2016; Hunnisett et al., 2024).

Metrics. We adopt standard CSP metrics and report both *sample-* and *crystal-level* scores:

1. *Collision rate* (Col_S): Fraction of generated samples with any intermolecular distance $< r_w - 0.7$ Å where r_w is the sum of atomic van der Waals radii (Cordero et al., 2008). Lower is better.
2. *Lattice match rate* (Lat_S and Lat_C): Using CSD COMPACK, a sample *matches* if at least 8 of 15 molecules can be aligned to the experimental cluster (cluster size per CSP5; see Appendix C). Lat_C is the fraction of targets with at least one match.
3. *Conformer recovery* (Rec_S and Rec_C): $\text{RMSD}_1 < 0.5$ Å (non-hydrogen) to a solid-state conformer. Rec_S averages over all samples; Rec_C is the fraction of targets with at least one match.
4. *Approximately solved* ($\widetilde{\text{Sol}}_C$): *Any* collision-free, lattice-matching sample with $\text{RMSD}_{15} < 2$ Å on a 15-molecule cluster.

4.1 CSP for Rigid and Flexible Molecules

First, we compare OXTAL to *ab initio* ML models on two different test sets comprised of representative rigid and flexible molecules with ground-truth structures in CSD (see §C.2 for details). For each crystal target, every method generates $n_S = 30$ samples. Training-time exclusions ensure

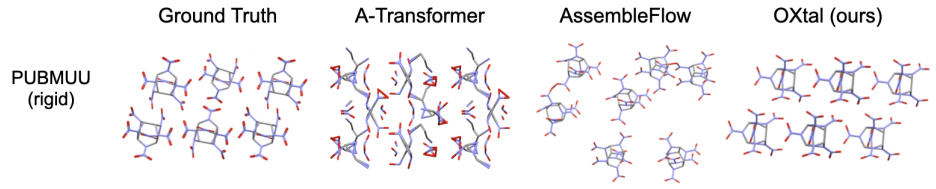


Figure 4: Example crystal packing generated by A-Transformer, AssembleFlow, and OXTAL.

Table 1: Performance of *ab-initio* machine learning models on both rigid and flexible molecular CSP. OXTAL achieves an order of magnitude improvement and is the only model able to approximately solve any crystals in the flexible dataset.

Model	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
			Rigid Dataset			
A-Transformer	0.731	0.015	0.060	0.033	0.120	0.060
AssembleFlow	0.524	0.001	0.040	0.001	0.020	0
OXTAL	**0.011**	**0.873**	**1.000**	**0.737**	**0.960**	**0.300**
			Flexible Dataset			
A-Transformer	0.874	0.002	0.063	0	0	0
AssembleFlow	0.850	0	0	0	0	0
OXTAL	**0.167**	**0.410**	**0.813**	**0.056**	**0.500**	**0.125**

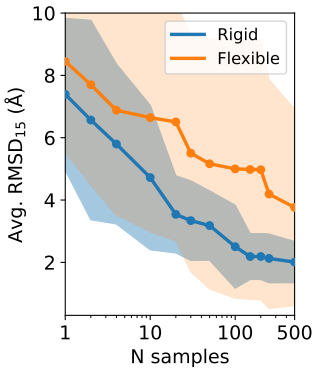


Figure 5: OXTAL sample efficiency for 10 rigid & flexible molecules.

no target appears in the training sets of OXTAL or A-Transformer; AssembleFlow uses a public checkpoint that may include overlap. DFT baselines are omitted here due to prohibitive cost.

Results. Table 1 shows OXTAL outperforms existing ML methods across all metrics on both datasets. Intramolecularly (Rec_C), OXTAL recovers up to 90% of solid-state molecular conformers; intermolecularly, Col_S is near zero on rigid targets and low on flexible ones. OXTAL's predicted packings also attain strong lattice match rate against experimental structures, with approximate solves for both rigid and flexible molecules. Qualitatively (Figure 4), A-Transformer struggles to capture meaningful conformers despite being given Z, highlighting the limitations of a unit cell based model. AssembleFlow generates molecular assemblies with large spatial separations that lack periodicity (also reflected in low Lat_C and Lat_S scores), along with frequent interatomic clashes.

Sample efficiency. For downstream screening and design, few-sample success is critical. OXTAL exhibits a log-linear improvement in RMSD_{15} among lattice-matched predictions as n increases (Figure 5), with several rigid targets reaching Lat_C and $\widetilde{\text{Sol}}_C$ with $n < 10$. This suggests the sampler (i) often lands near the correct motif and (ii) refines global periodicity with additional draws.

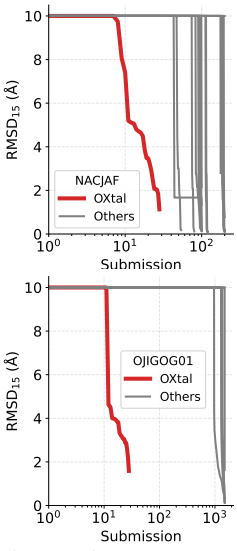


Figure 6: OXTAL, for targets shown, achieves similar RMSD_{15} with less (%) submissions compared to DFT.

4.2 CCDC CSP BLIND TESTS

Every few years, CCDC holds a CSP blind test competition, which invites leading computational chemistry groups to solve a handful of hidden crystal structures (Bardwell et al., 2011; Reilly et al., 2016; Hunnisett et al., 2024). We therefore evaluate OXTAL on structures from the three most recent (5th, 6th, and 7th) blind tests. Metrics and experiment set up follow §4.1. We compare OXTAL and other ML *ab initio* baselines against the aggregate of reported expensive DFT-based submissions (DFT_{avg}). See Appendix C.4 for details and per-structure results.

Results. Table 2 shows that OXTAL strongly outperforms *ab initio* ML baselines, and achieves the best or second best performance across all three tests using only 30 samples per target. While DFT methods may sometimes attain higher conformer recovery or approximate solve rates, when OXTAL is allowed to generate the same number of samples as the DFT methods for CSD Blind Test 5, it matches the approximate-solve

Table 2: Results for the 5th, 6th, and 7th CCDC CSP blind tests. Classical chemistry methods are aggregated as DFT_{avg}. The best model is **bolded**, and the second best is underlined.

Model	n_S	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
CSP Blind Test 5							
A-Transformer	180	0	0	0	0	0	0
AssembleFlow	180	0.717	0	0	0.150	0.500	0
DFT_{avg}	3314	**0.003**	0.307	0.556	**0.772**	0.681	**0.500**
OXTAL	180	0.006	**0.667**	**0.833**	0.572	**0.833**	0.167
CSP Blind Test 6							
A-Transformer	150	0	0	0	0	0	0
AssembleFlow	150	0.800	0	0	0.073	0.200	0
DFT_{avg}	1591	0.082	0.230	0.416	**0.490**	0.440	**0.400**
OXTAL	150	**0.013**	**0.660**	**1.000**	0.160	**0.600**	0.200
CSP Blind Test 7							
A-Transformer	240	0	0	0	0	0	0
AssembleFlow	240	0.808	0	0	0.063	0.250	0
DFT_{avg}	6608	0.067	0.053	0.500	**0.319**	**0.456**	**0.441**
OXTAL	240	**0.021**	**0.483**	**0.875**	0.129	0.375	0.125

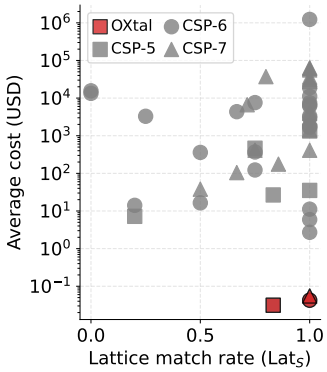


Figure 7: Lattice match rate per crystal attempted relative to average inference cost (in $USD) for submitted CCDC competition methods. OXTAL is denoted in red. Costs are normalized to a single on-demand AWS instance from Sept. 2025 (see §C.4.1).

rate of DFT_{avg} (see Appendix F.2). OXTAL consistently scores the best in terms of lattice match rate: from a per-sample basis, only $5 - 30\%$ of DFT samples match the lattice cluster, whereas OXTAL generally recapitulates the packing structure ($48 - 67\%$). This suggests that while DFT identifies many local energetic minima, OXTAL can better capture the joint energy and kinetic features that determine which minima are more likely to be formed. Of the thousands of submitted predictions (and potentially more generated but unsubmitted structures), only a small percentage of DFT-identified minima are close to ground truth structures (Figure 6). Predicting the correct experimental structure in few shots is crucial for downstream discovery applications, and OXTAL reliably approximates lattice packings *sans* any ranking methods.

Inference cost. Another major pitfall of DFT-based methods is the extremely high computational cost required to run the atomistic simulations. Recently, CCDC raised concerns over the 46 million CPU core hours that were reportedly utilized by submitted methods in CSP-7 to solve only 8 crystals (Hunnisett et al., 2024). Unlike traditional DFT methods that require new simulations for each new molecule, OXTAL's upfront training cost (§B) is amortized at inference, allowing us to efficiently generate samples for new molecules. Using standardized on-demand cloud pricing (§C.4.1), OXTAL *is over an order of magnitude cheaper at inference time*, while still achieving strong lattice match rates (Figure 7). Given the CCDC's call for more efficient CSP algorithms, OXTAL's cost profile enables broad screening and dense posterior sampling before any optional physics-based refinement.

4.3 Survey of Chemical Interpretability

Beyond benchmark metrics, we examine OXTAL's ability to reproduce chemically meaningful intra- and intermolecular features in practically relevant crystals. Figure 1 highlights accurate packings ($\text{RMSD}_{15} < 1.5$Å) across diverse rigid and flexible chemotypes: drug-like molecules (HURYUQ), polymer precursors (CAPRYL), organometallics (ACACPD), π-conjugated materials (QQQCIG,), and QAXMEH, which contains the most known polymorphs (see §H for more).

Molecular interactions. For intramolecular interactions, OXTAL recovers *solid-state* geometries for highly flexible molecules (which are highly influenced by packing), including small-molecule drugs as well as biomolecular fragments in the Protein DataBank (e.g. $\text{RMSD}_1 = 1.3$Å for a 6-mer peptide with 17 rotatable bonds (2OKZ) in Figure 8(a)). For intermolecular interactions, OXTAL accurately captures both strong and weak interactions, both in registry and lengths. Examples in Figure 8(b) include the complementary hydrogen bonds in a semiconducting crystal (XIJJOT), the weak Cl-H halogen bonds in an organometallic catalyst (OJIGOG), π-π stacking in π-functional molecule (TEPNIT), and multiple weak contacts in a cluster of the flexible drug aripiprazole (MELFIT) in Figure 18. Zooming out, while OXTAL may somewhat prefer co-planar motifs, it can still reproduces diverse packing motifs, including 1D columnar structures (e.g. BALNAD in

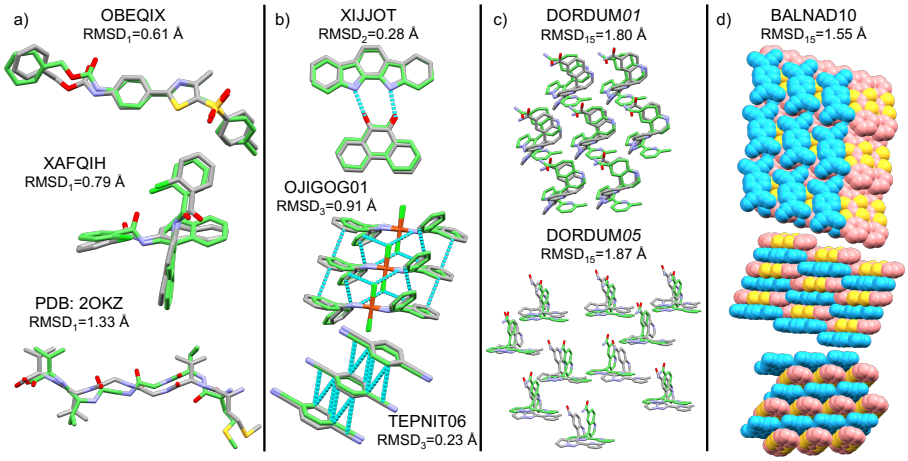


Figure 8: OXTAL (green) captures experimental (grey) (a) intramolecular and (b) intermolecular interactions in drug-like molecules, peptides, semiconductors, and catalysts. OXTAL can further infer (c) distinct experimental polymorphs as well as (d) co-crystals.

Figure 8(d)), quasi-2D herringbone structures (e.g. ANTCEN in Figure 16), and 2D brickwork structures (e.g. UMIMIO in Figure 1).

Polymorphs. For molecules with many known polymorphs (including highly rotatable scaffolds), independent OXTAL samples predict distinct experimental polymorphs (e.g., the drugs galunisertib DORDUM in Figure 8(c) and indomethacin INDMET in Figure 18). This suggests the sampler can *explore multiple kinetic and thermodynamic basins* rather than collapsing to a single motif. The diversity of sampled polymorphs are shown in Figure 19.

Multi-component systems. Lastly, OXTAL is not restricted to single-component systems. OXTAL can correctly predict the interactions between electron donor and acceptors (Figures 8 and 17), which dictate the crystals' electronic properties. For the charge-transfer semiconducting co-crystals BALNAD and PERTCQ, OXTAL correctly reproduces the 1D π-stacked columns with alternating donors and acceptors, the precise inter-columnar brick-wall registry, as well as the π-π stacking distances.

5 Related Works

Physical approaches to crystal structure prediction. Classical crystal structure prediction (CSP) mostly applied search-based algorithms to sampled a pre-defined search space (Hunnisett et al., 2024). While many such methods have shown varied degree of success for different applications (van Eijck, 2002; Pickard & Needs, 2011; Case et al., 2016; Tom et al., 2020; Banerjee et al., 2021), they fundamentally rely on many calls to an expensive evaluation function (e.g., energy computation using DFT) and struggle to leverage prior data effectively. Recent search and optimization approaches replace DFT with machine learning interatomic potentials (Batatia et al., 2024; Wood et al., 2025; Gharakhanyan et al., 2025). In contrast, OXTAL does not require explicit energy function calls and brings orders of magnitude improvements in speed and inference costs.

Generative models for inorganic crystal structure prediction. Generative models have been applied for unconditional de-novo generation of new inorganic periodic crystals (Xie et al., 2022) and later used for generation conditioned on crystal composition (Jiao et al., 2023; 2024; Yang et al., 2024; Miller et al., 2024; Levy et al., 2025). In contrast to molecular crystals, inorganic crystals are typically smaller and do not possess the same flexibility/packing diversity.

Protein structure prediction. In computational structural biology, the analogous task of protein structure prediction (PSP) conditioned on a protein sequence has seen transformative progress with landmark models like AlphaFold (Jumper et al., 2021; Abramson et al., 2024) and ESMFold (Lin et al., 2023), followed by de novo protein design approaches (Watson et al., 2023; Yim et al., 2023; Bose et al., 2024). These models work with a few dozen residue types, compared to a larger chemical space in general molecular CSP, and use MSA and evolutionary structural information not present in molecular crystals. Detailed discussion of related work can be found in §G.

6 DISCUSSION

In this paper, we introduce OXTAL, a large-scale all-atom diffusion model for 3D molecular CSP that learns the joint distribution of molecular conformations and periodic packing conditioned on 2D graphs. Discarding explicit equivariance and unit-cell parametrization in favor of a symmetry-aware S^4 augmentation, OXTAL learns periodic motifs from locally consistent neighborhoods at scale, enabling efficient sampling at all-atom resolution. Empirically, OXTAL achieves state-of-the-art results among *ab initio* ML methods, and attains competitive lattice recovery at several orders of magnitude lower cost compared to DFT-based methods. Our chemical survey supports OXTAL's ability to capture diverse intra- and intermolecular interactions, yet several improvements remain. These include integrating reliable ranking and local relaxation, conditioning on crystallization context (i.e. solvent, temperature), and further improving sample efficiency and uncertainty quantification.

ETHICS STATEMENT

This work models experimentally realizable molecular crystal structures using publicly curated crystallographic data (CSD) under an academic license; we redistribute only non-proprietary data and provide scripts that require users to obtain their own CSD access. No human subjects or personally identifiable information are involved. Potential impacts are largely beneficial (accelerated solid-form screening and molecular materials design), but we acknowledge dual-use risks: faster structure generation might marginally aid the design of hazardous energetic materials or hard-to-detect polymorphs that affect drug bioavailability. We mitigate by (i) releasing usage guidelines and filters to flag energetic motifs and highly strained lattices, and (ii) encouraging downstream human/physics validation before deployment in safety-critical settings. We report training compute and approximate carbon footprint in §B; we adopt mixed precision and efficient dataloading to reduce energy use. The authors declare no competing financial interests.

REPRODUCIBILITY STATEMENT

We will release code, model checkpoints, and evaluation scripts. Because CSD redistribution is restricted, the training code will not be released, but appropriate pre-processing code, such as S^4 cropping (including r_{cut}, r^0, $T_{\max}$, $K_{\max}$) will be released. We fix random seeds for training and sampling, and specify all hyperparameters necessary for reproducing the model. We will specify third-party tool versions (RDKit, xTB/GFN2, COMPACK). Evaluation is fully scripted: COMPACK settings, $\text{RMSD}_1/_{15}$ definitions, collision thresholds ($r_w - 0.7$ Å), and criteria for $\text{Lat}_\text{S}/\text{Lat}_\text{C}$, $\text{Rec}_\text{S}/\text{Rec}_\text{C}$, and $\widetilde{\text{Sol}}_C$. We will also provide inference configurations for all reported figures/tables, ablation toggles, baseline re-runs (A-Transformer, AssembleFlow) with their seeds, and a Docker container to reproduce numbers on comparable hardware.

REFERENCES

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. (Cited on pages 2, 5, 6, 9, 23, and 27)

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. (Cited on page 4)

Luis M Antunes, Keith T Butler, and Ricardo Grau-Crespo. Crystal structure generation with autoregressive large language modeling. *Nature Communications*, 15(1):10570, 2024. (Cited on page 34)

Atreyee Banerjee, Dipti Jasrasaria, Samuel P Niblett, and David J Wales. Crystal structure prediction for benzene using basin-hopping global optimization. *The Journal of Physical Chemistry A*, 125 (17):3776–3784, 2021. (Cited on pages 9 and 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. *Structural Science*, 67(6):535–551, 2011. (Cited on pages 1, 6, 7, 22, 24, and 25)

Luis Barroso-Luque, Muhammed Shuaibi, Xiang Fu, Brandon M Wood, Misko Dzamba, Meng Gao, Ammar Rizvi, C Lawrence Zitnick, and Zachary W Ulissi. Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771*, 2024. (Cited on page 34)

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *NeurIPS*, 2022. (Cited on page 34)

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhowmik, Filippo Bigi, Samuel M. Blau, Vlad Cărare, Michele Ceriotti, Sanggyu Chong, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Fabio Falcioni, Edvin Fako, Andrea C. Ferrari, John L. A. Gardner, Mikolaj J. Gawkowski, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Jonas Grandel, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Christian Holm, Cheuk Hin Ho, Stephan Hofmann, Jad Jaafar, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Panagiotis Kourtis, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Chen Lin, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Louise A. M. Rosset, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Viktor Svahn, Christopher Sutton, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Santiago Vargas, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Thomas Wolf, Fabian Zills, and Gábor Csányi. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2024. (Cited on pages 9 and 34)

Avishek Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se(3)-stochastic flow matching for protein backbone generation, 2024. URL https://arxiv.org/abs/2310.02391. (Cited on page 9)

Carl Ivar Branden and John Tooze. *Introduction to protein structure*. Garland Science, 2012. (Cited on page 2)

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. (Cited on pages 9 and 34)

Jie Chen, Bipul Sarma, James MB Evans, and Allan S Myerson. Pharmaceutical crystallization. *Crystal growth & design*, 11(4):887–895, 2011. (Cited on page 1)

Aleksandr Aleksandrovich Chernov. *Modern crystallography III: crystal growth*, volume 36. Springer Science & Business Media, 2012. (Cited on page 1)

Beatriz Cordero, Verónica Gómez, Ana E Platero-Prats, Marc Revés, Jorge Echeverría, Eduard Cremades, Flavia Barragán, and Santiago Alvarez. Covalent radii revisited. *Dalton Transactions*, (21):2832–2838, 2008. (Cited on page 6)

Farren Curtis, Xiayue Li, Timothy Rose, Alvaro Vazquez-Mayagoitia, Saswata Bhattacharya, Luca M Ghiringhelli, and Noa Marom. Gator: a first-principles genetic algorithm for molecular crystal structure prediction. *Journal of chemical theory and computation*, 14(4):2246–2264, 2018. (Cited on page 34)

Qianggang Ding, Santiago Miret, and Bang Liu. Matexpert: Decomposing materials discovery by mimicking human experts. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=AUBvo4sxVL. (Cited on page 34)

David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005. (Cited on page 34)

Eberhard Engel and Reiner M Dreizler. *Density functional theory*. Springer, 2011. (Cited on page 1)

P Ganguly and Gautam R Desiraju. Long-range synthon aufbau modules (lsam) in crystal structures: systematic changes in c 6 h 6- n f n ($0 \leq n \leq 6$) fluorobenzenes. *CrystEngComm*, 12(3):817–833, 2010. (Cited on page 34)

Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–6802, 2021. (Cited on pages 4 and 34)

Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ulissi, C Lawrence Zitnick, and Abhishek Das. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. *arXiv preprint arXiv:2204.02782*, 2022. (Cited on page 34)

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. (Cited on pages 9 and 34)

Hongyu Guo, Yoshua Bengio, and Shengchao Liu. Assembleflow: Rigid flow matching with inertial frames for molecular assembly. In *ICLR*, 2025. (Cited on pages 6, 23, and 24)

Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky T. Q. Chen, and Yaron Lipman. Generator matching: Generative modeling with arbitrary markov processes. In *International Conference on Machine Learning*, 2025. (Cited on page 4)

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), 2024. (Cited on pages 1, 6, 7, 8, 9, 24, and 25)

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=DNdN26m2Jk. (Cited on pages 9 and 34)

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. URL https://openreview.net/forum?id=jkvZ7v4OmP. (Cited on pages 9 and 34)

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. (Cited on pages 2, 6, and 9)

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. (Cited on page 6)

Daniel S Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G Taylor, Muhammad R Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, et al. The open molecules 2025 (omol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762*, 2025. (Cited on page 34)

Daniel Levy, Siba Smarak Panigrahi, Sékou-Oumar Kaba, Qiang Zhu, Kin Long Kelvin Lee, Mikhail Galkin, Santiago Miret, and Siamak Ravanbakhsh. SymmCD: Symmetry-preserving crystal generation with diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xnssGv9rpW. (Cited on pages 9 and 34)

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *ICLR*, 2024. (Cited on pages 6, 28, and 34)

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. (Cited on page 9)

Andriy O Lyakhov, Artem R Oganov, Harold T Stokes, and Qiang Zhu. New developments in evolutionary structure prediction algorithm uspex. *Computer Physics Communications*, 184(4): 1172–1182, 2013. (Cited on page 34)

Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. lddt: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, 2013. (Cited on page 21)

Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023. (Cited on page 34)

Benjamin Kurt Miller, Ricky TQ Chen, Anuroop Sriram, and Brandon M Wood. Flowmm: Generating materials with riemannian flow matching. In *ICML*, 2024. (Cited on pages 4, 9, and 34)

Aaron J Nessler, Okimasa Okada, Mitchell J Hermon, Hiroomi Nagata, and Michael J Schnieders. Progressive alignment of crystals: reproducible and efficient assessment of crystal structure similarity. *Applied Crystallography*, 55(6):1528–1537, 2022. (Cited on page 22)

Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pp. 65–84. Springer, 2003. (Cited on page 4)

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023. (Cited on page 6)

Chris J Pickard and RJ Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011. (Cited on pages 9 and 34)

Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22(14): 7169–7192, 2020. (Cited on page 5)

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. *Structural Science*, 72(4): 439–459, 2016. (Cited on pages 1, 6, 7, 24, and 25)

Luis Reinaudi, Ezequiel PM Leiva, and Raúl E Carbonio. Simulated annealing prediction of the crystal structure of ternary inorganic compounds using symmetry restrictions. *Journal of the Chemical Society, Dalton Transactions*, (23):4258–4262, 2000. (Cited on page 34)

Nate Schultheiss and Ann Newman. Pharmaceutical cocrystals and their physicochemical properties. *Crystal growth and design*, 9(6):2950–2967, 2009. (Cited on page 1)

Justin S Smith, Roman Zubatyuk, Benjamin Nebgen, Nicholas Lubbers, Kipton Barros, Adrian E Roitberg, Olexandr Isayev, and Sergei Tretiak. The ani-1ccx and ani-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data*, 7(1):134, 2020. (Cited on page 34)

Anuroop Sriram, Sihoon Choi, Xiaohan Yu, Logan M Brabson, Abhishek Das, Zachary Ulissi, Matt Uyttendaele, Andrew J Medford, and David S Sholl. The open dac 2023 dataset and challenges for sorbent discovery in direct air capture, 2024. (Cited on page 34)

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, et al. Protenix-advancing structure prediction through a comprehensive alphafold3 reproduction. *BioRxiv*, pp. 2025–01, 2025. (Cited on pages 20 and 21)

Rithwik Tom, Timothy Rose, Imanuel Bier, Harriet O'Brien, Álvaro Vázquez-Mayagoitia, and Noa Marom. Genarris 2.0: A random structure generator for molecular crystals. *Computer Physics Communications*, 250:107170, 2020. (Cited on pages 9 and 34)

Bouke P van Eijck. Crystal structure predictions for disordered halobenzenes. *Physical Chemistry Chemical Physics*, 4(19):4789–4794, 2002. (Cited on pages 9 and 34)

Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Crystal structure prediction via particle-swarm optimization. *Physical Review B—Condensed Matter and Materials Physics*, 82(9):094116, 2010. (Cited on page 34)

Yu Wang, Lingjie Sun, Cong Wang, Fangxu Yang, Xiaochen Ren, Xiaotao Zhang, Huanli Dong, and Wenping Hu. Organic crystalline materials in flexible electronics. *Chemical Society Reviews*, 48 (6):1492–1530, 2019. (Cited on page 1)

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. (Cited on page 9)

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. (Cited on pages 9 and 34)

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. URL https://openreview.net/forum?id=03RLpj-tc_. (Cited on pages 9 and 34)

Sherry Yang, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. Scalable diffusion for materials generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=wm4WlHoXpC. (Cited on page 9)

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se(3) diffusion model with application to protein backbone generation. *International Conference on Machine Learning (ICML)*, 2023. (Cited on page 9)

Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. Graphsaint: Graph sampling based inductive learning method. *arXiv preprint arXiv:1907.04931*, 2019. (Cited on page 5)

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. (Cited on page 2)

Xiaotao Zhang, Huanli Dong, and Wenping Hu. Organic semiconductor single crystals for electronics and photonics. *Advanced Materials*, 30(44):1801048, 2018. (Cited on page 1)

APPENDIX

A THEORY

Proposition 1. *Let $\partial \mathbf{A}_{crop} = \{\{u, v\} \in E : u \in \mathbf{A}_{crop}, v \notin \mathbf{A}_{crop}\}$ represent the boundary of $\mathbf{A}_{crop}$. Denote the number of atoms in a volume C as $T(C)$. Let $L_\partial(\mathbf{A}_{crop}) = \sum_{\{u,v\}\in\partial\mathbf{A}_{crop}} \mathcal{L}(u, v)$ be the boundary loss. Assuming $\exists r^0$ s.t. $L(u, v) = 0, \forall u, v$ s.t. $\|u-v\| > r^0$, i.e. $\mathcal{L}$ is local, and there exist $0 < a \leq b < \infty$ s.t. $a|S| \leq T(S) \leq b|S|$ for any $S \subseteq V$. Then,*

$$\frac{L_\partial(\mathbf{A}_{crop})}{T(\mathbf{A}_{crop})} = O((1+\epsilon)T(\mathbf{A}_{crop})^{-1/3}) \tag{6}$$

To prove this proposition, we first label our assumptions.

A1 **Locality.** $\exists r_0$ s.t. $L(u, v) = 0$ for all u, v s.t. $\|u - v\| > r_0$

A2 **Uniform Density.** there exist $0 < a \leq b < \infty$ s.t. $a|S| \leq T(S) \leq b|S|$ for any $S \subseteq V$.

Next we investigate the S4 algorithm. We first recall the definition of shells and $\mathbf{A}_{crop}$

$$\mathcal{S}_k(g_c) = \{g_i \in \mathcal{C}^{(U)} : k\epsilon \leq d(g_c, g_i) < (k+1)\epsilon\}, \quad k = 0, 1, 2, \ldots \tag{8}$$

where d is the distance between molecules g_c and g_i in the infinitely repeating lattice. Molecules from each full shell are added until the total number of tokens satisfies $|V_{g_c}| + \sum_{k\in\mathcal{K}} |V_{\mathcal{S}_k}| \leq N_{\max}$, where $N_{\max}$ is the crop budget. If at least one full shell fits, we train on the crop $\mathbf{A}_{crop} = \{g_c\} \cup \bigcup_{k\in\mathcal{K}} \mathcal{S}_k$. This implies that $\mathbf{A}_{crop}$ can be well approximated by a ball centered around g_c or radius $(k+1)\epsilon$.

Define the ball of radius ϵk as $\mathcal{B}_k = \bigcup_{k\in\mathcal{K}} \mathcal{S}_k$. First a short lemma to show the number of neighbors of a node is bounded.

Lemma 1. *Define the **edge neighborhood** of g as $\mathcal{N}(g) = \{\{u, v\}$ s.t. $u = g \cap \mathcal{L}(u, v) \geq 0\}$. Assuming [A1] and [A2] we have a uniform bound on the degree*

$$|\mathcal{N}(g)| \leq C \tag{9}$$

for some C independent of g.

Proof. This follows from considering a ball of radius r_0 around g. That ball has volume $V = \frac{4}{3}r_0^3$. Using [A2] we have that the token count $T(\mathcal{B}_k(g)) \leq \frac{4b}{3}r_0^3$, and therefore $|\mathcal{N}(g)| < \frac{4b}{3}r_0^3$ which is independent of g. □

We next note that for a regular lattice structure we have the following inequality for the size of the boundary relative to the total volume. Specifically for a lattice we have

Lemma 2. *For $0 < a \leq b < \infty$ on a 3D lattice we have the following relationship between a sphere's surface area and volume $a|\mathcal{B}_k|^{2/3} \leq |\partial\mathcal{B}_k| \leq b|\mathcal{B}_k|^{2/3}$.*

we note that the constants are due to discretization error, and locality. As we approach the continuous limit i.e. $k \to \infty$ the bounds become tight.

Next, we have to bound how well $\mathbf{A}_{\text{crop}}$ is approximated by a ball of radius k.

Lemma 3. *For some constant $0 < c < \infty$, $\partial\mathbf{A}_{crop} \leq \partial\mathcal{B}_k + c\epsilon|\mathcal{B}_k|^{2/3}$.*

Proof. We first note that $\mathcal{B}_k \subseteq \mathbf{A}_{\text{crop}}\mathcal{B}_{k+1}$ by construction of $\mathbf{A}_{\text{crop}}$. This means that we can consider $\mathbf{A}_{\text{crop}}$ as all the molecules in $\mathcal{B}_k$ plus (possibly) some additional molecules in $\mathcal{S}_k(g_c)$. We can then bound the total surface area as

$$|\partial\mathbf{A}_{\text{crop}}| \leq |\partial\mathcal{B}_k| + |\partial\mathcal{S}_k(g_c)| \tag{10}$$

however we know that the surface area of molecules in $\mathcal{S}_k(g_c)$ is bounded by the number of nodes in $\mathcal{S}_k$ and some constant.

$$|\partial\mathcal{S}_k(g_c)| \leq c|\mathcal{S}_k| \tag{11}$$
$$= c((\epsilon(k+1))^3 - (\epsilon k)^3)) \tag{12}$$
$$= c\epsilon^3(3k^2 + 3k + 1) \tag{13}$$
$$\leq c\epsilon^3 k^2 \tag{14}$$
$$\leq c\epsilon|\mathcal{B}_k|^{2/3} \tag{15}$$

which combined with equation 10 proves the lemma. □

We are now ready to prove the main proposition.

Proof. Assuming $\mathcal{L}_\partial(\mathbf{A}_{crop})$ is based on a sum of pairwise interaction terms i.e.

$$\mathcal{L}_\partial(\mathbf{A}_{crop}) = \sum_{\{u,v\}\in\partial\mathbf{A}_{crop}} \mathcal{L}(u,v) \tag{16}$$

Lemma 3 allows us to first analyze $\mathcal{L}_\partial(\mathcal{B}_k)$ then use Lemma 3 for the final bound.

$$\mathcal{L}_\partial(\mathcal{B}_k) = \sum_{\{u,v\}\in\partial\mathcal{B}_k} \mathcal{L}(u,v) \tag{17}$$
$$\leq c|\partial\mathcal{B}_k| \max_{\{u,v\}\in\partial\mathcal{B}_k} \mathcal{L}(u,v) \tag{18}$$

using Lemma 2,

$$\leq c|\mathcal{B}_k|^{2/3} \max_{\{u,v\}\in\partial\mathcal{B}_k} \mathcal{L}(u,v) \tag{19}$$

Assuming [A1], [A2], and Lemma 1,

$$\leq c|T(\mathcal{B}_k)|^{2/3} \tag{20}$$

Combining this with Lemma 3 we have the final result

$$\frac{L_\partial(\mathbf{A}_{\text{crop}})}{T(\mathbf{A}_{\text{crop}})} = O((1+\epsilon)T(\mathbf{A}_{\text{crop}})^{-1/3})$$

□

As a reminder, this proposition implies that the boundary surface becomes less of an issue as the number of tokens grows.

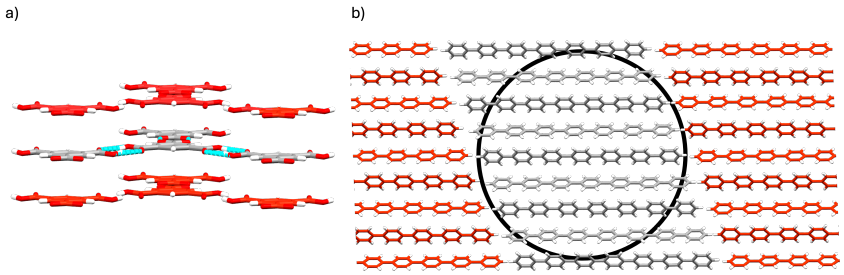


Figure 9: (a) KNN as used in AlphaFold3 captures only the closest interactions (e.g. hydrogen bonds in trimesic acid, highlighted in blue) but does not capture more distant interactions that are equally crucial for crystallization (e.g. π-π stacking in trimesic acid, highlighted in red). (b) centroid-based approaches will create anisotropic crops in elongated molecules, for example only capturing a 1-dimensional column (grey) in the ellipsoid *p*-quinquephenyl, while missing peripheral interactions (red).

B Technical Details

Experiments were performed on heterogenous GPU clusters consisting of NVIDIA L40S and H100 GPUs. Models were primarily trained using multinode DDP training on L40S GPUs as cluster availability permitted. Inference is performed across all available hardware.

B.1 Training Data Processing

We curate training data from the Cambridge Structural Database (CSD, including releases up to May 1, 2025) under the following criteria: (i) 3D coordinates are present, (ii) the conventional R-factor $< 9\%$, (iii) the structure is derived from single crystal diffraction at ambient pressure, (iv) the structure is not polymeric, (v) the structure can be sanitized by RDKit with no missing heavy atoms, (vi) there is a known space group, and (vii) at most 250 non-hydrogen atoms are present per unit cell. To avoid leakage, we exclude any entry belonging to a test crystal family and any entry containing a molecular component that appears in test sets. To avoid oversampling certain clusters, within each crystal family, near-duplicate polymorphs are collapsed using $\mathrm{RMSD}_{15} \leq 0.25$ Å, retaining the entry with the lowest R. Our final training dataset contains 594,202 crystals in total.

For data preparation, crystallographic disorder is resolved by selecting the disorder group with the highest occupancy. We remove hydrogens for training and extract kekulized bonds from crystallographic metadata. Prior to building supercells, molecules of the crystals are centered to lie within the unit cell, and the unit cells are transformed to their unique Niggli-reduced forms. Supercells are constructed by tile translation $\mathbf{T}_{ijk} = i\mathbf{a} + j\mathbf{b} + k\mathbf{c}$ for $i, j, k \in -1, 0, 1$ such that molecules with centroids inside the supercell boundary are included.

B.2 Additional Model Details

B.2.1 Stoichiometric Stochastic Shell Sampling (S^4)

We provide the full algorithm for S^4 cropping in Algorithm 1. Recall that for a crystal $\mathcal{C}$, $\mathcal{M}$ denotes the set of molecules within the crystal, $\mathbf{A}$ denotes the tokenized atom array for all molecules $m \in \mathcal{M}$, and $X \subset \mathbb{R}^3$ denotes the Cartesian coordinates of each atom $a \in \mathbf{A}$. Note that for practical purposes, we assume $\mathcal{M}$ has a finite size. Additionally, $\mathcal{A}_\mathcal{C}$ is the crystal's minimal *asymmetric unit*, and d is a pre-computed intermolecular distance matrix, where $d(i, j) = \min_{x_i \in X(m_i), x_j \in X(m_j)} ||x_i - x_j||_2$.

Given an input shell radius r_{cut}, we first sample a central molecule $m_c \sim \mathrm{Uniform}(\mathcal{A})$. Then, we assign all other molecules $m_i \in \mathcal{M} \setminus m_c$ to a shell layer, as defined by r_{cut}. Note that each molecule m_i can only belong to one shell layer. Next, with probability $(1 - p_{\max})$, we randomly sample how many shells to keep. We then add complete shells of molecules to our selected set until the maximum token budget $T_{\max}$ is reached. If no complete shell can fit within the token budget, we adaptively sample molecules in the first shell according to Algorithm 2 in order to preserve the stoichiometric ratios of the molecules in $\mathcal{A}_\mathcal{C}$. Lastly, we note that it is possible to deduce the underlying Bravais lattice of a large periodic point cloud by analyzing the set of interatomic difference vectors (a Patterson

analysis): in the resulting Patterson function, lattice translation vectors appear as a high-multiplicity, regularly spaced subset of the interatomic vectors, from which the lattice parameters (and hence a primitive cell) can be recovered.

Implementation details. In practice, we begin by first considering a $3 \times 3 \times 3$ crystal supercell $\mathcal{C}^{(U)}$, where $U = 3I_3$. For each input crystal, we sample a molecule m_c from the asymmetric unit of the central unit cell. After analysing the distribution of intermolecular distances for all crystals in the training set, we decided to set $r_{cut} = 4.5$ Å (see Figure 10). Furthermore, in order to encourage the selection of larger $\mathbf{A}_{crop}$ blocks, we set $p_{\max} = 0.8$ and $T_{\max} = 640$.

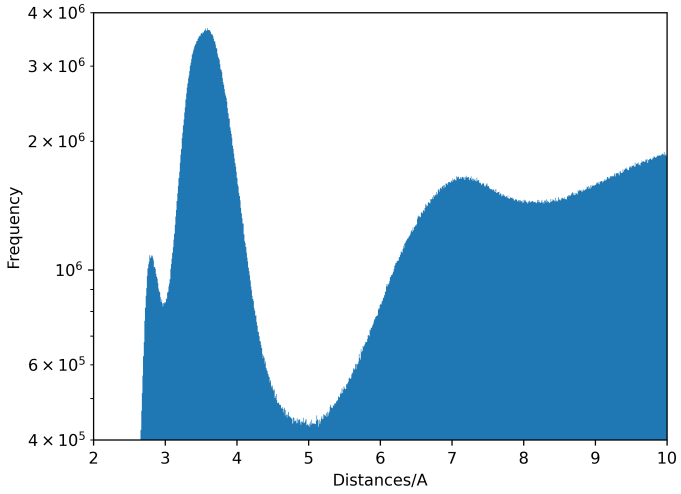


Figure 10: Distribution of intermolecular distances in the processed CSD training dataset.

Algorithm 1 STOICHIOMETRIC STOCHASTIC SHELL SAMPLING

Input: $\mathcal{M}$, $\mathbf{A}$, $\mathcal{A}$, d, r_{cut}, T_{max}, p_{max}
$m_c \sim \text{Uniform}(\mathcal{A})$ ▷ Sample central molecule
$\mathcal{M} \leftarrow \mathcal{M} \setminus \{m_c\}$
Initialize $\mathcal{S} = \emptyset$, $k \leftarrow 1$
while $\mathcal{M} \neq \emptyset$ **do** ▷ Compute shell layers around m_c
  $S_k \leftarrow \{m_i \in \mathcal{M} : d(m_c, m_i) \leq k \cdot r_{cut}\}$
  $\mathcal{S} \leftarrow \mathcal{S} \cup S_k$, $\mathcal{M} \leftarrow \mathcal{M} \setminus S_k$, $k \leftarrow k + 1$
end while

$b_{\text{max}} \sim \text{Bernoulli}(p_{\text{max}})$ ▷ Sample maximum number of shells $S_{k_{\max}}$ to keep
if $b_{\text{max}} = \text{False}$ **then**
  $k_{\max} \sim \text{Uniform}\{1, \ldots, k-1\}$
  $\mathcal{S} \leftarrow \{S_1, \ldots, S_{k_{\max}}\}$
end if

Initialize $\mathbf{A}_{\text{crop}} \leftarrow \mathbf{A}(m_c)$, $i \leftarrow 1$,
while $|\mathbf{A}_{\text{crop}}| \leq T_{\max}$ **do** ▷ Add full integer shells within token budget $T_{\max}$
  if $|\mathbf{A}_{\text{crop}}| + |\mathbf{A}(S_i)| > T_{\max}$ **then**
    break
  end if
  $\mathbf{A}_{\text{crop}} \leftarrow \mathbf{A}_{\text{crop}} \cup \mathbf{A}(S_i)$, $i \leftarrow i + 1$
end while

if $i = 1$ **then** ▷ If no full shell fits, sample molecules according to stoichiometry
  $\mathbf{A}_{\text{crop}} \leftarrow$ ADAPTIVE STOICHIOMETRIC SAMPLING($\mathbf{A}$, m_c, $\mathcal{A}$, T_{max}, S_1)
end if
return $\mathbf{A}_{\text{crop}}$

Algorithm 2 ADAPTIVE STOICHIOMETRIC SAMPLING

Input: $\mathbf{A}$, m_c, $\mathcal{A}$, T_{max}, S
Calculate proportion of each molecule type $p_t = \frac{|\{m_i \in \mathcal{A} : \text{type}(m_i) = t\}|}{|\mathcal{A}|}$
Initialize ideal targets: $R_t \leftarrow \text{round}(p_t \cdot |S|)$ for each type t
$\mathbf{A}_{\text{crop}} \leftarrow \mathbf{A}(m_c)$, $R_{\text{type}(m_c)} \leftarrow R_{\text{type}(m_c)} - 1$ ▷ m_c is already pre-selected

while $|\mathbf{A}_{\text{crop}}| < T_{\max}$ **do**
  Bucket remaining molecules by type: $B_t = \{m \in S : \text{type}(m) = t\}$
  Compute adaptive weights for each type:
  for all $t \in T$ with $B_t \neq \emptyset$ **do**
    $w_t \leftarrow R_t / |B_t|$
  end for
  Normalize weights to probabilities: $P'_t = w_t / \sum_t w_t$
  Select type t' randomly according to probabilities P'_t
  Sample $m' \sim \text{Uniform}(B_{t'})$
  if $|\mathbf{A}_{\text{crop}}| + |\mathbf{A}(m')| > T_{\max}$ **then**
    break ▷ Stop if adding this molecule exceeds token budget
  end if
  $\mathbf{A}_{\text{crop}} \leftarrow \mathbf{A}_{crop} \cup \mathbf{A}(m')$
  $S \leftarrow S \setminus \{m'\}$, $B_{t'} \leftarrow B_{t'} \setminus \{m'\}$, $R_{t'} \leftarrow R_{t'} - 1$
end while
return $\mathbf{A}_{crop}$

B.2.2 ARCHITECTURE

Our model adopts the AlphaFold-3 (AF3) trunk-plus-diffusion design via the public PROTENIX Team et al. (2025) implementation. Atom-level tokens with relative position and entity encodings

are processed by an AF3-style *Pairformer* trunk with recycling to produce single (s) and pair (z) representations; the MSA and LM/RNA pathways are disabled in all reported experiments. These representations condition a structure-denoising head consisting of an atom-attention encoder, a transformer-based diffusion module, and an atom-attention decoder that outputs 3D coordinates. Aside from retokenizing residues as atoms and disabling MSA/LM inputs, we make no material architectural changes relative to AF3.

B.2.3 LOSSES

We use a combination of losses to encourage accurate structure prediction on both the global and local scales. Here we enumerate what the different losses are, how they are formulated, what they are useful for and how they lead to a correct diffusion model.

SmoothLDDTLoss. The Smooth local distance difference test (SmoothLDDT) loss is a smooth form of an existing LDDT loss (Mariani et al., 2013). The LDDT measures how well two structures align over all pairs of atoms at a distance closer than some predefined threshold. For us, this is $R_0 = 15\mathring{A}$. These pairs of atoms form a set of local distances L. The LDDT score is the average of fractions at four distance thresholds [0.5, 1.0, 2.0, 4.0] Angstroms. The Smooth LDDT test, instead of using a binary test of whether or not the local distances are on the same side of the threshold, uses a sigmoid function instead. Let

$$L = \|x - x^T\| \tag{21}$$

$$L^{GT} = \|x^{GT} - (x^{GT})^T\| \tag{22}$$

$$\delta = \text{abs}(L - L^{GT}) \tag{23}$$

$$\epsilon = \frac{1}{4}\left[\text{sigmoid}(\frac{1}{2} - \delta) + \text{sigmoid}(1 - \delta) + \text{sigmoid}(2 - \delta) + \text{sigmoid}(4 - \delta)\right] \tag{24}$$

$$\text{mask} = \delta < 15 \tag{25}$$

Then the SMOOTHLDDT between two molecules of size d is:

$$\text{SMOOTHLDDT}(x, x^{GT}) = \sum(\epsilon \cdot \text{mask})/(d(d-1)) \tag{26}$$

Our SmoothLDDT loss is then equal to

$$\mathcal{L}_{\text{sLDDT}}(x, x^{GT}) = \text{SMOOTHLDDT}(x, \text{ALIGN}(x^{\text{GT}}, x)) \tag{27}$$

where $\text{ALIGN}(a, b)$ performs an optimal rigid alignment of a to b in 3D.

B.2.4 MODEL HYPERPARAMETERS

We use the following hyperparameters for training and inference.

Training. We base our training implementation off of the open-source Protenix model (Team et al., 2025). Because we are not doing protein folding, we have disabled the previously built-in multiple sequence alignment (MSA) module. We train using an Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.95$, and a learning rate of 0.0018 and weight decay of 1e-8. The learning rate is additionally set to a scheduler with 1,000 warmup steps, and a decay factor of 0.95 for every 50,000 steps. We report results for our model trained at 110,000 steps.

The atom encoder has 3 blocks with 4 heads, and the atom embedding size is 128, whereas the pair embedding is size is 16. In terms of the main pairformer trunk, we use 4 pairformer blocks, each with 16 attention heads and a dropout rate of 0.25. The hidden dimension for the pair representation is 128, and the hidden dimension for the single representations are 384. Furthermore, we do 2 rounds of recycling. For the diffusion block, we use a diffusion batch size of 32 and a chunk size of 1. The actual diffusion transformer has a 12 transformer blocks, with 8 attention heads. As mentioned above, crop size (and therefore token size for the diffusion transformer) is set to 640.

Inference. At inference time, we use the following hyperparameters for our diffusion sampling: $\gamma_0 = 0.8$, noise scale lambda = 1.003, and we use 200 steps with a step scale eta of 1.5. When generating samples, we generate 1 sample from 30 different seeds for each target crystal structure.

C Experimental Details

C.1 Selection of Performance Metrics

The four metrics in the main text jointly probe physical plausibility (Col_S), packing similarity (Lat_S, Lat_C), intramolecular fidelity (Rec_S, Rec_C), and a holistic approximate solve indicator ($\widetilde{\text{Sol}}_C$). Reporting both sample-level and crystal-level aggregates is essential: the former characterizes the distributional quality of all proposals, while the latter answers whether a target was recovered at least once. Concretely, crystal-level rates are an OR-aggregation over a target's samples. The pair helps diagnose failure modes such as mode collapse (high Lat_C with low Lat_S) and low recall (the reverse).

Thresholds and interpretability. We adopt community conventions for RMSD_k (Nessler et al., 2022):

- $\text{RMSD}_{15} < 1.0$ **Å** typically indicates the predicted packing reproduce the experimental match and lies in the exact energy basin as the experimental structure. This metric is the one used by the 5th CSP blind test (Bardwell et al., 2011).
- $1.0-2.0$ **Å** usually indicates the prediction shares the correct topology with mild lattice strain or small reorientations. If the H-bond graph, Z/Z', density, and PXRD are consistent, this is very likely recoverable to < 1 Å with a brief local relaxation. Otherwise, the structure is often structurally related to the ground truth (i.e. in or near the correct packing motif/topology or space group but with slip, molecular misorientation, or a cell mismatch). It demonstrates that the method can correctly identify the neighborhood of the global energy minimum, even if it can't pinpoint the exact minimum itself.
- $2.0-3.0$ **Å** typically indicate a significant mismatch. At this level of deviation, key intermolecular interactions, like hydrogen bonds, may be incorrect. The overall packing symmetry might also be different. However, the prediction might still capture a general feature of the packing (e.g., identifying a layered structure vs. a herringbone packing).
- **above** 3.0 **Å**. These values are generally not considered useful. The deviation is so large that the predicted structure is almost certainly in a completely different and incorrect energy basin. Any similarity to the experimental structure is likely coincidental.

Our $\widetilde{\text{Sol}}_C$ (collision-free, lattice-matching, $\text{RMSD}_{15} < 2$ Å) is a thus a strict, early-stage -utility indicator: not "solved," but reliably near-correct for downstream relaxation and re-ranking.

Lattice matching for non-periodic predictions.. Our generators output finite blocks without periodic conditions. For large inference blocks, it is possible to identify translation vectors that map the finite cluster onto itself (e.g., by correlating molecular centroids/local environments), least-squares fit three independent vectors to define a primitive lattice, and convert atoms to fractional coordinates. Nevertheless, CSD *COMPACK* can be directly used for robust alignment of a block against a crystal structure while preserving standard CSP rigor. First, we employ standard practices of avoiding hydrogen-atom alignment and allowing mismatches in connectivity annotations (e.g., bond orders). In the absence of periodic images, greedy pruning can miss valid correspondences; therefore for all methods (including DFT baselines) we use a search time of 10 seconds with a distance/angle threshold of 50%. This enables reliable COMPACK outputs on non-PBC inputs *without diluting the criterion*: acceptance still hinges on multi-molecule superposition (15-molecule cluster, ≥ 8 matched) and the same RMSD_k thresholds used in periodic CSP benchmarks.

C.2 Molecular CSP on Rigid and Flexible Datasets

We construct our rigid and flexible datasets using crystals from the Cambridge Crystal Structure Database (CCDC), following the same processing procedure outlined in §B.1. The rigid dataset is comprised of 50 molecular crystals, and the flexible dataset is comprised of 16 molecular crystals. These are generally crystals with practical relevance or newly-released crystals. We generally define rigid molecules to contain 0-3 rotatable bonds, and flexible molecules to contain more than that. We note there is a small nuance in flexibility, as we refine it in the molecular context (by ring restriction or number of known polymorphs, etc.), this means rubrene (QQQCIG), tetraphenylporphryin (TPHPOR), CL-20 (PUBMUU) are defined as rigid; ROY (QAXMEH), galunisertib (DORDUM), sulfathizaole (SUTHAZ), flufenamic acid (FPAMCA) and YOPYEL are defined as 'flexible'. The exact CSD identifiers are provided below for reference.

Table 3: Performance of *ab-initio* machine learning models on rigid and flexible molecular CSP. For each model, results are calculated using $n = 30$ samples for each crystal target.

Model	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
	Rigid Dataset					
A-Transformer	0.731	0.015	0.060	0.033	0.120	0.060
AssembleFlow	0.524	0.001	0.040	0.001	0.020	0
AlphaFold3	0.114	0.089	0.340	0.133	0.480	0
OXTAL	0.011	0.873	1.000	0.737	0.960	0.300
	Flexible Dataset					
A-Transformer	0.874	0.002	0.063	0	0	0
AssembleFlow	0.850	0	0	0	0	0
AlphaFold3	0.580	0.144	0.313	0.191	0.4375	0
OXTAL	0.167	0.410	0.813	0.056	0.500	0.125

Molecules in the rigid dataset: XULDUD, GUFJOG, QAMTAZ, BOQQUT, BOQWIN, UJIRIO, PAHYON, XATMIP, HAMTIZ, XATMOV, AXOSOW, SOXLEX, WIDBAO, HXACAN, ACSALA, PUBMUU, GLYCIN, TEPNIT, QQQCIG, ZZZMUC, FOYNEO, ANTCEN, URACIL, BTCOAC, CORONE, GUJTOX, TPHPOR, ACETAC, IMAZOL, CEBYUD, CILJIQ, CUMJOJ, DEZDUH, DOHFEM, GACGAU, GOLHIB, HURYUQ, IHEPUG, LECZOL, NICOAM, ROHBUL, UMIMIO, WEXREY, CAPRYL, ACACPD, BPYRUF, JIVNOV, MUBXAM, VUJBUB, NAVZAO.

Molecules in the flexible dataset: QAXMEH, DORDUM, BOTHUR, MOVZUW, YOPYEL, SUTHAZ, TPPRHC, FPAMCA, INDMET, MELFIT, YIGPIO, TEHZIP, DMANTL, YIGDUP, UWEQUL, COWZIA.

C.3 *Ab initio* Machine Learning Baselines

We compare OXTAL against three *ab initio* machine learning baselines: (i) a lightweight transformer trained with flow matching on our dataset (§C.3.1), (ii) the publicly released AssembleFlow-Atom model (Guo et al., 2025), and (iii) AlphaFold3 (Abramson et al., 2024) used as a generative baseline. Our complete table of results are reported in Table 3 on both rigid and flexible molecular CSP benchmarks using $n = 30$ samples per crystal target.

C.3.1 A-Transformer (ours)

Model. A lightweight transformer encoder operating on atom tokens with unit-cell features. The model is a PyTorch `nn.TransformerEncoder` with hidden size $d_h = 512$, $L = 13$ layers, and $H = 4$ heads. Each atom embedding includes element/charge, time $t \in (0, 1]$, fingerprints, and, when enabled, unit-cell lengths (a, b, c) and angles (α, β, γ). Two output heads are used: a coordinate head ($\mathbb{R}^3$ per token) and, optionally, a rotation head (unit quaternion, disabled in our main runs).

Training. The objective is rigid-cluster translation flow matching in $\mathbb{R}^3$. We linearly interpolate between ground-truth translations s_0 and random in-cell starts s_1,

$$s_t = (1 - t)\, s_0 + t\, s_1, \quad t \sim \text{Unif}[t_{\min}, 1],$$

and predict denoised coordinates $\hat{x}_0$. The loss is an x_0 regression term:

$$\mathcal{L} = \lambda_{\text{trans}} \sum_i \|x_0^{(i)} - \hat{x}_0^{(i)}\|^2.$$

Inference and Evaluation. At inference we integrate from t=1 to $t_{\min}$ with Euler steps, producing clusters in a P1 box (no lattice recovery). Metrics are RMSD_1 and RMSD_k only; lattice metrics are not applicable.

This baseline is deliberately minimal: it ignores rotation flow matching and lattice prediction, relying solely on translation flow matching and unit-cell features. It provides a capacity-matched reference against which to measure the benefits of OXTAL's architecture and training.

C.3.2 AssembleFlow

We evaluate the released AssembleFlow-Atom model (Guo et al., 2025) using the authors' checkpoints without re-training or fine-tuning. For each molecule we generate clusters of 17 rigid copies from an RDKit ETKDG conformer, applying random rotations and translations with uniform offsets $[-S, S]^3$, where $S \in \{10, 15, 20\}$ Å. Three seeds $\{42, 7, 2024\}$ and seven checkpoints yield 63 runs per molecule. To standardize evaluation, we randomly sample 30 outputs per crystal and compute metrics on those.

Inference follows the official `position_inference` routine with default settings. Final coordinates are wrapped into a large P1 box (no lattice recovery). We report best-of-grid RMSD_k.

AssembleFlow enforces rigid-molecule assembly and provides a strong baseline for rigid-packing quality, but it does not predict lattice parameters and therefore cannot be scored on lattice metrics.

C.3.3 AlphaFold3

We additionally evaluate AlphaFold3 as a structural generative baseline. For each target molecule, we generate conformations using default protein–ligand generative settings, treating each conformer as a candidate crystal packing. Inference is performed based on 30 counts of the molecule, per our other methods. Thirty samples per target are evaluated in the same way as for other baselines. While AlphaFold3 was not designed for crystal structure prediction, including it highlights the gap between general-purpose biomolecular structure generators and CSP-specific models.

C.4 CCDC CSP Blind Test Details

To contextualize our work, we provide a summary of the 5th, 6th, and 7th CCDC CSP blind tests. These community-wide challenges have benchmarked the state-of-the-art in predicting the crystal structures of organic molecules from their chemical graphs alone, primarily relying on density functional theory (DFT). We here provide brief descriptions of these tests, and refer the audience to (Bardwell et al., 2011; Reilly et al., 2016; Hunnisett et al., 2024) for details. Performance is assessed using the COMPACK algorithm, which quantifies structural similarity through root mean square deviation (RMSD) of molecular clusters.

In the first four CSP blind tests, the field evolved from force-field landscapes to the first widespread use of periodic dispersion-corrected DFT (DFT-D), which proved decisive in 4th blind test and set the stage for DFT to become the de facto standard for final lattice-energy evaluation. The fifth blind test (2010) marked a significant increase in the complexity of the target molecules. The six targets included rigid molecules, semi-flexible molecules, a 1:1 salt, a highly flexible pharmaceutical-like compound, and two co-crystals. This test highlighted what has become a standard workflow in CSP: broad structure generation (e.g., grid/quasi-random sampling, parallel tempering), followed by local minimization and hierarchical re-ranking. While at least one successful prediction was submitted for each target, the success rate was lower than in previous tests, underscoring the increased difficulty. DFT-D demonstrated its reliability for discriminating between competing structures for small and moderately flexible molecules, but handling high flexibility and complex solid forms remained a major challenge.

The sixth blind test (2015/16) continued with five challenging targets: a small, nearly rigid molecule; a polymorphic former drug candidate; a chloride salt hydrate; a co-crystal; and a large, flexible molecule. This test solidifed the "search–optimize–rank" pipeline. On the search side, more than half of the methods allowed intramolecular flexibility during exploration, and many adopted hierarchical filtering starting with generating conformer and packing, followed by progressively tighter optimization and pruning. In optimization and ranking, dispersion-corrected periodic DFT became mainstream, with vdW models (e.g., D3/D3(BJ), MBD) and multipole-based electrostatics or SAPT-derived potentials used to refine close competitors. All experimental structures were predicted by at least one submission except a potentially disordered Z' = 2 polymorph. The results demonstrated that while DFT-D provided reliable baseline energetics, accurate treatment of conformational flexibility remained the primary bottleneck.

The seventh test (2022-2022) introduced a two-phase structure to separate structure generation from ranking challenges. The seven targets featured a silicon-iodine containing molecule, a copper coordination complex, a near-rigid molecule, a co-crystal, a polymorphic small agrochemical, a highly flexible polymorphic drug candidate, and a polymorphic morpholine salt. The test also featured

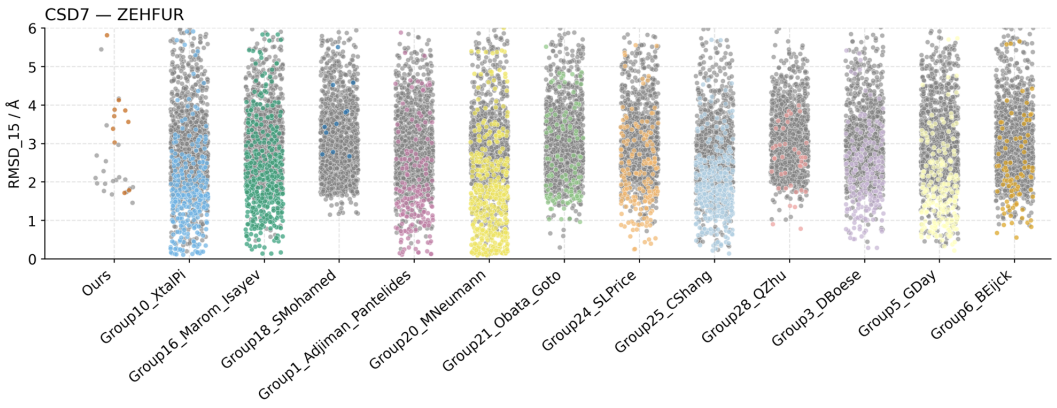


Figure 11: Beeswarm plot of submitted structures for Molecule XXXI (ZEHFUR) from CSD7. In 30 inference samples, OXTAL obtains performance comparable to some DFT groups with thousands of submitted samples. Colored points signify a partial lattice match whereas gray points signify lattice mismatch (e.g. 3 out of 15 molecules match and have low RMSD, but the packing is significantly different).

one of the most challenging systems in the history of the blind tests: a large, highly polymorphic pharmaceutical drug candidate. A key finding from the structure generation phase was that while different search methods often identified overlapping sets of low-energy structures, their success rates tended to decrease as the complexity of the target molecule increased. For ranking, periodic dispersion-corrected GGA DFT methods produced results in excellent agreement with experiment across most targets, whereas higher-level corrections, full free-energy treatments, and ML on DFT potentials were less decisive for these specific systems than expected. Additionally, dynamic/static disorder (and high Z') remained difficult.

These blind tests shows a clear trend that DFT-based ranking has become reliable, but limitations persist: computational expense, difficulty capturing kinetic effects, challenges with disorder and flexibility and high Z' number, and the fundamental limitation that thermodynamic ranking often fails to predict which polymorphs are experimentally observable. Our approach aims to address these limitations by directly learning both thermodynamic and kinetic regularities from data, with the hope of eventually eliminating the need for extensive search, local optimization, and post-hoc ranking altogether. By generating a small number of high-quality structures directly, we bypass the traditional generate-optimize-rank pipeline that has dominated CSP. Below tables (Table 9, Table 10, and Table 11) in §I show comparisons for all submitted structures by each group (sans ranking). The beeswarm plot (Figure 11) of molecular XXXI of test submission further highlights the extensive sampling currently required by DFT methods compared to our method. This point is further highlighted in Figure 12, which plot the best RMSD_1 or RMSD_{15} achieved at a given n submission size for a few blind test crystals.

Because each group can pick and choose which crystals they want to solve for, we report an aggregated metric in §4.2 of all submitted DFT results. For crystal-level metrics, we consider the total number of crystal targets across all DFT methods to be $n_{\text{groups}} * n_{\text{targets}}$. We then compute aggregate metrics - counting groups who did not submit any submissions for a crystal as a miss, since they technically did not solve it. For sample-level metrics, we simlpy compute the average across all submitted structures.

C.4.1 INFERENCE TIME COST

We computed computational costs by multiplying the reported wall–clock compute time by an AWS on–demand unit price with hours taken directly from the three blind tests references (Bardwell et al., 2011; Reilly et al., 2016; Hunnisett et al., 2024). Because the papers report heterogeneous processors and often normalize times to ∼3.0 GHz core–hours, we treat one normalized CPU hour as one AWS vCPU–hour; we then price CPU time at the c5.large on–demand rate in Sept. 2025, i.e., \$0.085 per *2* vCPUs ⇒ \$0.0425 per vCPU–hour, so CPU cost = hours × \$0.0425. For L40S GPU runs, we map directly to `g6e.xlarge` (which contains one NVIDIA L40S) and price at \$1.861 per instance–hour, so GPU cost = hours × \$1.861. We apply the rates pro–rata without further rounding.

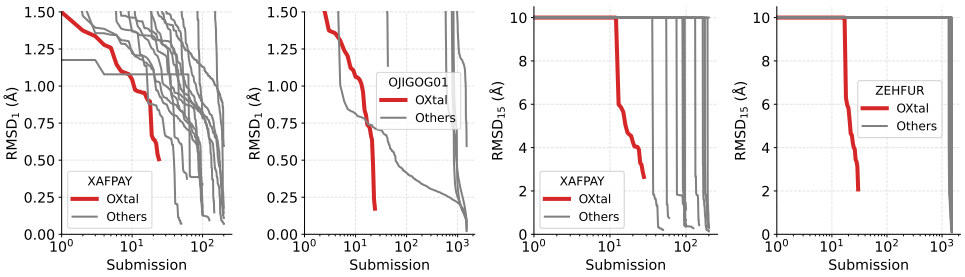


Figure 12: Sample efficiency plots XAFPAY (blind test 6, a flexible molecule), OJIGOG01 (blind test 7), and ZEHFUR (blind test 7). OXTAL is able to quickly recover both molecular conformers ($RMSD_1$) and periodicity ($RMSD_{15}$).

All prices are on–demand EC2 instances as of September 2025. For reference, exact reported times used are presented in Table 12, Table 13, and Table 14.

D Model Ablation Studies

Here, we provide a series of different model ablation studies. Due to the high computational demand of training OXTAL, all in this section are reported at 50k training steps instead of the 110k reported in the main body of the paper. Following our predefined evaluation protocol, we evaluate $n_S = 30$ generated samples with 30 molecular copies for each crystal target.

D.1 Cropping Method Ablation

In order to isolate the performance gain from S^4, we directly compare it against other standard cropping methods. Overall, S^4 performs the best across all metrics and datasets with centroid radius performing the second best, indicating that there may not be many oblong molecules in the evaluation datasets. However, these results support the benefits of using S^4, since it is not subject to the same pitfalls of KNN and centroid radius cropping as illustrated in Figure 9.

Also, note that our implementation of centroid radius cropping is more generous than the standard one used in the literature (Abramson et al., 2024). Specifically, naive radius cropping selects any atom that lies within a pre-defined spatial radius. However, this often results in partial molecules being cropped, which naturally leads to lower performance. Instead, we consider all atoms in a given molecule if its *centroid* lies within 15 Å of a randomly sampled central molecule, and only consider complete molecules up to our token budget, which mitigates the issue of fragmented molecules during training. KNN is also implemented such that it only considers complete molecules up to the token budget, ordered by neighbor distance from a randomly sampled central molecule.

Table 4: Results for different cropping methods with best results **bolded** and second best underlined.

Crop Method	Dataset	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
KNN	Rigid	0.086	0.631	0.920	0.520	0.780	0.220
	Flexible	0.580	0.164	0.688	0.005	0.062	0
	CSP5	0.085	0.470	0.667	0.433	0.833	0
	CSP6	0.132	0.353	0.800	0.007	0.200	**0.200**
	CSP7	0.312	0.165	**0.750**	0.005	0.125	0
Centroid Radius	Rigid	0.040	0.669	**0.980**	0.591	**0.940**	**0.340**
	Flexible	0.425	0.301	0.688	0.048	0.375	0.125
	CSP5	0.054	0.470	**0.833**	0.429	0.833	0
	CSP6	0.151	0.324	**1.000**	**0.173**	0.400	0
	CSP7	0.104	**0.225**	**0.750**	**0.104**	0.250	0
Ours (S^4)	Rigid	**0.026**	**0.688**	0.940	**0.629**	**0.940**	0.280
	Flexible	**0.302**	**0.346**	**0.750**	**0.050**	**0.438**	**0.188**
	CSP5	**0.000**	**0.500**	**0.833**	**0.478**	**1.000**	0
	CSP6	**0.047**	**0.440**	**1.000**	0.107	**0.800**	**0.200**
	CSP7	**0.033**	0.221	0.625	**0.104**	**0.500**	0

D.2 S^4 Radius Size Ablation

Next, we investigate the sensitivity of OXTAL to the choice of S^4 shell radius size. From our results in Table 5, we see that there are generally minimal changes in performance across three different choices of radius size. Recall that our main OXTAL model uses a shell radius size of 4.5, which was selected from analyzing the distribution of intermolecular distances presented in Figure 10. From this figure, both 4.5 and 5 Å are natural choices for the first shell, however due to the "layered" nature of crystallization, we consider a second shell at 9 rather than 10 Å to be more fitting.

Overall, our current implementation of S^4 may slightly favor smaller shell radius sizes due to the overall token budget constraint, which limits the number of full shells we are able to accommodate. Regardless, we see that all three radius settings for S^4 outperform existing cropping methods from Table 4, suggesting that OXTAL is somewhat robust to reasonable choices in S^4 radius size.

Table 5: Effect of S^4 shell radius size, with best results **bolded** and second best underlined.

Radius Size	Dataset	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
3.5 Å	Rigid	**0.015**	**0.694**	**0.980**	0.596	0.920	**0.320**
	Flexible	**0.275**	**0.394**	0.688	0.041	0.375	**0.250**
	CSP5	0.006	0.439	0.833	0.433	0.833	0
	CSP6	0.068	0.409	**1.000**	**0.174**	0.600	**0.400**
	CSP7	0.095	**0.238**	**0.750**	0.110	**0.500**	0
4.5 Å	Rigid	0.026	0.688	0.940	**0.629**	0.940	0.280
	Flexible	0.302	0.346	**0.750**	**0.050**	**0.438**	0.188
	CSP5	**0.000**	**0.500**	0.833	**0.478**	**1.000**	0
	CSP6	**0.047**	**0.440**	**1.000**	0.107	**0.800**	0.200
	CSP7	**0.033**	0.221	0.625	0.104	**0.500**	0
5.5 Å	Rigid	0.047	0.690	0.940	0.625	**0.960**	0.220
	Flexible	0.414	0.346	**0.750**	0.028	0.312	0.188
	CSP5	0.012	0.470	0.667	0.439	0.667	0
	CSP6	0.147	0.375	**1.000**	0.147	0.600	0.200
	CSP7	0.101	0.234	**0.750**	**0.142**	0.375	0

D.3 Equivariance Ablation

To highlight the importance of model scaling that a non-equivariant architecture provides, we perform an ablation study with EquiformerV2 (Liao et al., 2024). In order to accommodate the larger molecular crystals in our training dataset, we are forced to significantly reduce the parameter budget for EquiformerV2 to fit it into memory. Specifically, we reduce the batch size to 4, and use a hidden dimension size of 32, 2 layers, and 4 attention heads. The rest of the training pipeline, including the use of S^4 cropping, remain the same as for OXTAL. Overall, the EquiformerV2 version of the model has a total parameter size of 15M, as opposed to the 100M total parameter size of OXTAL.

Table 6: Performance of the explicitly equivariant EquiformerV2 model compared to OXTAL.

Model	Dataset	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
EquiformerV2	Rigid	0.044	0.009	0.272	0.009	0.060	0
	Flexible	**0.084**	0	0	0	0	0
	CSP5	0.006	0	0	0	0	0
	CSP6	0.181	0	0	0	0	0
	CSP7	0.267	0	0	0	0	0
OXTAL	Rigid	**0.026**	**0.688**	**0.940**	**0.629**	**0.940**	**0.280**
	Flexible	0.302	**0.346**	**0.750**	**0.050**	**0.438**	**0.188**
	CSP5	**0.000**	**0.500**	**0.833**	**0.478**	**1.000**	0
	CSP6	**0.047**	**0.440**	**1.000**	**0.107**	**0.800**	**0.200**
	CSP7	**0.033**	**0.221**	**0.625**	**0.104**	**0.500**	0

From Table 6, we see that EquiformerV2 performs quite well on the collision metric, indicating that the model is generally able to produce physically feasible structures. Furthermore, it is also able to recover a few lattice packings and molecular conformers from the Rigid molecule dataset. However, EquiformerV2 completely fails to recover any lattices or conformers from the more challenging Flexible and CSP Blind Test datasets, which is unsurprising due to its drastically smaller parameter budget.

These results show the necessity of using a scalable, non-equivariant model: enforcing equivariance imposes a critical bottleneck on model size, which renders it insufficient for capturing larger molecular crystals in the complex crystallization landscape. This validates our design choice of scaling via a non-equivariant model and breaking free from lattice constraints while softly enforcing symmetry constraints.

D.4 Model Size Ablation

To further investigate the effect of model size, we train a smaller version of OXTAL that contains roughly 50M total parameters, which is half the size of the one we report in the main paper.

Table 7: Effect of OXTAL model size, with best results in **bold** and second best underlined.

Model	Dataset	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
A-Transformer	Rigid	0.731	0.015	0.060	0.033	0.120	0.060
	Flexible	0.874	0.002	0.063	0	0	0
	CSP5	0.833	0	0	0	0	0
	CSP6	0.967	0	0	0	0	0
	CSP7	0.950	0	0	0	0	0
AssembleFlow	Rigid	0.524	0.001	0.040	0.001	0.020	0
	Flexible	0.850	0	0	0	0	0
	CSP5	0.717	0	0	0.150	0.500	0
	CSP6	0.800	0	0	0.073	0.200	0
	CSP7	0.808	0	0	0.063	0.250	0
OXTAL (50M)	Rigid	0.040	0.247	0.680	0.213	0.680	0.160
	Flexible	0.423	0.253	0.562	0.018	0.188	0.062
	CSP5	0.018	0.164	0.500	0.200	0.833	0
	CSP6	0.077	0.253	0.800	0.066	0.400	**0.200**
	CSP7	0.076	0.193	**0.750**	0.097	0.250	**0.125**
OXTAL (100M)	Rigid	**0.026**	**0.688**	**0.940**	**0.629**	**0.940**	**0.280**
	Flexible	**0.302**	**0.346**	**0.750**	**0.050**	**0.438**	**0.188**
	CSP5	**0.000**	**0.500**	**0.833**	**0.478**	**1.000**	0
	CSP6	**0.047**	**0.440**	**1.000**	**0.107**	**0.800**	**0.200**
	CSP7	**0.033**	**0.221**	0.625	**0.104**	**0.500**	0

As shown in Table 7, halving the parameter budget of OXTAL significantly reduces performance. However the smaller 50M parameter OXTAL model still outperforms existing *ab-initio* ML methods, which we have provided again here for reference. This result reinforces the importance of scalability while also highlighting the benefits of OXTAL's lattice-free and non-equivariant architecture at a smaller scale.

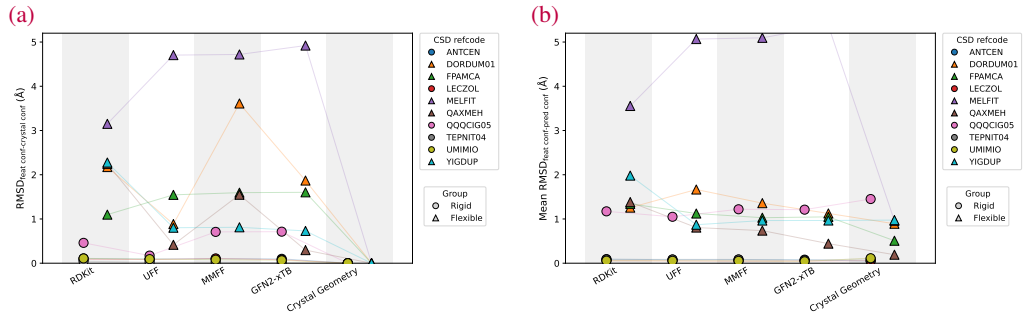


Figure 13: Analysis of input and final conformer in ten crystals for which OXTAL obtains $\text{RMSD}_1 < 0.5$Å. (a) Different input conditioning conformers and their differences with the ground truth. Generally, flexible molecules' conformers have significant deviations to the crystal geometry. (b) Different input conditioning conformers and their differences with the final, model-predicted conformer. Generally, the model final conformers that are significantly from the conditioning.

E CONFORMER ANALYSIS

We now investigate the influence of the feature molecular conformer used by OXTAL. For each molecule, the model is conditioned on (i) its bond information and (ii) a 3D structure obtained using RDKit ETKDG followed by relaxation with the semi-empirical quantum-chemical method GFN2-xTB. This conformer serves solely as a feature conditioning signal for the computational prior, rather than as an experimental oracle: the generative process always starts denoising from random atomic coordinates, with the conformer providing only a reference for physically plausible initialization derived from quantum-chemical approximations as a feature input. To quantify how this initialization influences generation quality, we perform the following targeted analyses:

1. Distance of the feature conformer relative to ground truth. In Figure 13(a), we show that the feature conditioning conformers for flexible molecules can be significantly different than the ground truth crystal conformation. This shows how often the model must depart from or refine the initial geometry.
2. Distance of the feature conformer relative to the prediction. In Figure 13(b), we show that there are significant differences between the feature conformer with the final, model-predicted conformer. Structures predicted by the diffusion process is different than the one conditioned in the input embedding, i.e. it is not merely reproducing the structure used in input embedding.
3. Robustness to feature conformer quality. We assess robustness by replacing the GFN2-xTB initialization with alternative or deliberately perturbed conformers,
   - RDKit-ETKDG conformers
   - Universal forcefield conformers
   - MMFF94 conformers
   - GFN2-xTB conformers
   - Ground truth crystal conformers.

   In Figure 14(b), we plot generation performance (e.g., RMSD_{15}) as a function of input distortion level. We also plot the best RMSD_{15} for each case in Figure 14(a). These plots show that the model's predicted crystal packing is not significantly affected by the different conformers or their differences with the ground truth crystal conformation.
4. The issue of Z'. When $Z' > 1$, the asymmetric unit contains several symmetry-independent molecules. These molecules can have similar conformation (but in different orientation) or genuinely different conformers. $\sim 10\%$ of the training set contains such examples, and in this subset, mean Z' is 2.14. OXTAL conditions every molecular copy with the same feature conformer, and can occasionally approximately solve crystals with $Z' > 1$ (e.g. target XXIII in CSD 6, XAFPAY02, $\text{RMSD}_{15} = 1.91$Å.)

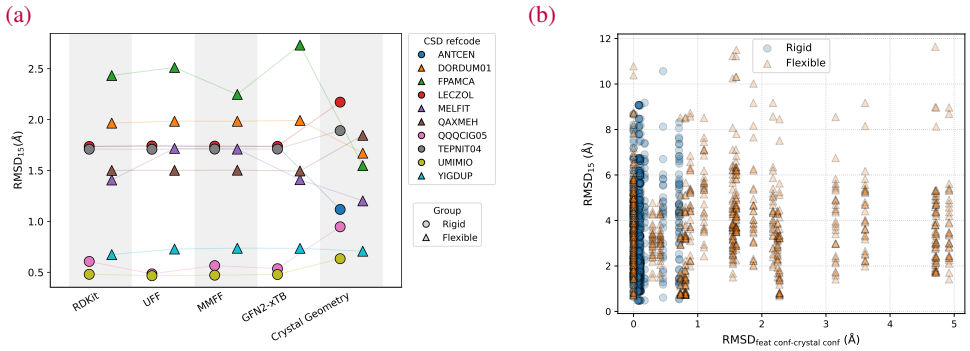


Figure 14: Analysis of input conformer with the final predicted crystal packing. (a) Best $RMSD_{15}$ from 30 samples per target when using different initial conformers. Generally, the model is robust against different sources of the input conditioning conformer. (b) All $RMSD_{15}$ across ten targets relative to the difference in input conditioning and crystal conformer differences. This shows the model can still produce good packings even if the input conditioning conformer there is very different from the crystal conformation

Together, these analyses provide a clear and quantitative view of how the initial conformer affects OXTAL: how different it is from the target, how effectively the model corrects it, how robust generation is to poor or perturbed inputs, and whether supplying multiple conformers per S^4 unit impacts performance.

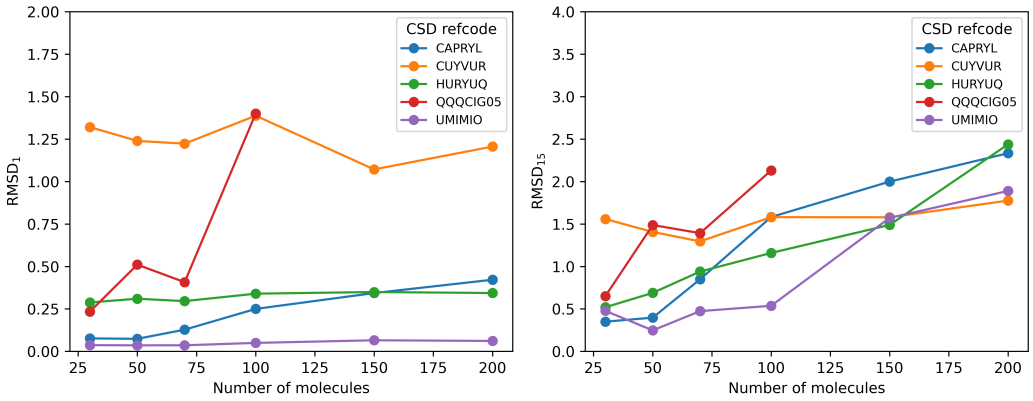


Figure 15: Performance of OXTAL on increasingly larger inference crystal blocks (denoted by the number of molecule copies generated). $RMSD_1$ (left) remains roughly constant, whereas $RMSD_{15}$ increases slightly for larger predicted packings. Note that QQQCIG05, which contains 42 heavy atoms (i.e. tokens) per molecules, goes out of memory past 100 molecules.

F Inference Analysis

In this section, we provide a deeper analysis into the inference and generation capabilities of OXTAL. Specifically, we investigate the performance of OXTAL on generating significantly larger crystal blocks, evaluate how the model performs when allowing for additional sampling, and provide a small qualitative analysis on the diversity of generated samples.

F.1 Inference on Larger Crystal Blocks

To evaluate the long-range periodicity of OXTAL, we perform inference on increasingly larger crystal blocks for a handful of different crystals (Figure 15). The respective number of atoms, and therefore tokens, for each molecule are provided in parentheses as follows: CAPRYL (10), CUYVUR (18), HURYUQ (14), QQQCIG05 (42), UMIMIO (12). Recall that OXTAL is trained with a maximum token budget of 640 (Appendix B.2.4), hence generating 200 copies of CAPRYL requires 2,000 tokens and 100 copies of QQQCIG05 requires 4,200 tokens, which are far beyond anything the training dataset would have seen.

In general, conformer recovery (denoted by $RMSD_1$) remains somewhat constant as more molecule copies are generated. This makes sense because if the model is already able to capture the correct crystal conformer, generating more copies of that should not change the conformer itself significantly. In terms of lattice periodicity (denoted by $RMSD_{15}$), OXTAL does struggle slightly with generating more molecules, since the packing arrangements cover distances that are further away. However, we note that although the $RMSD_{15}$ does increase for these larger blocks, they still mostly remain under the 2.0 Å threshold to be considered "approximately solved."

This analysis supports the long-range generalizability of OXTAL to generate larger periodic packings, and we provide a visualization of the approximately solved larger packing for ANTCENE in Figure 16, which contains over 2,400 tokens.

F.2 Additional Sample Evaluation for CSP Blind Test 5

For all previously reported results, we evaluate OXTAL using 30 generated samples per crystal target. This is done in order to highlight the sample efficiency of OXTAL (cf. Figures 5 and 6), since few-sample success is critical for downstream screening and design. However, we note that this puts OXTAL at somewhat of a disadvantage when comparing against traditional DFT methods, which often generate hundreds or thousands of candidate packings. Here, we provide a more direct comparison of OXTAL to DFT methods by evaluating OXTAL using the same number of generated samples on the CSP Blind Test 5 dataset.

As expected, allowing OXTAL to generate more candidate packings increases its performance on our per-crystal evaluation metrics, while remaining roughly the same on per-sample metrics. Notably, we also see that increasing the number of generated samples bring the approximate solve rate of

Table 8: Evaluation of OXTAL on CSP Blind Test 5 with the same number of generated samples per crystal target ($\overline{n_S}$) as DFT_{avg}.

Model	$\overline{n_S}$	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
DFT_{avg}	500	**0.003**	0.307	0.556	**0.772**	0.681	**0.500**
OXTAL	30	0.006	**0.667**	0.833	0.572	0.833	0.167
OXTAL	500	0.009	0.536	**1.000**	0.548	**1.000**	**0.500**

OXTAL equal to that of DFT_{avg}. This suggests that the model sampler is indeed able to generate accurate crystal packings and compete with traditional DFT methods.

G Additional Related Work

Physical approaches to crystal structure prediction. Physical approaches mostly rely on search and sampling using an energy function. Some classical methods infused domain knowledge, such as starting with initial guesses, such as a unit cell, and varying parameters random sampling (Case et al., 2016; Pickard & Needs, 2011; Tom et al., 2020), guidance from force-fields (van Eijck, 2002), or constructed guesses based on chemical principles (Ganguly & Desiraju, 2010). Recent methods have also applied more structured search algorithm, such as simulated annealing (Reinaudi et al., 2000; Earl & Deem, 2005), genetic algorithm (Curtis et al., 2018; Lyakhov et al., 2013), particle swarm optimization (Wang et al., 2010) and basin-hopping (Banerjee et al., 2021).

ML potentials. Computational complexity of DFT and availability of large datasets (Levine et al., 2025; Sriram et al., 2024; Barroso-Luque et al., 2024; Smith et al., 2020) enabled the development of universal machine learning interatomic potentials (MLIP) (Gasteiger et al., 2021; 2022; Batatia et al., 2022; Liao et al., 2024; Batatia et al., 2024; Wood et al., 2025) to predict energy and forces of different atomistic systems (from small molecules to inorganic crystals) at a fraction of DFT costs. The next generation of physical approaches to CSP replaced DFT with MLIPs and showed benefits in material discovery (Merchant et al., 2023) with recent FastCSP (Gharakhanyan et al., 2025) applied to organic CSP.

Generative models for inorganic crystal structure prediction. Generative models have emerged as a promising new paradigm for crystal structure prediction focusing mainly on conformer search in molecular structures as well as inorganic, periodic crystals. For inorganic crystal structures, which consist of a periodic unit cell of atoms, generative models were first applied for unconditional de-novo generation of new crystal (Xie et al., 2022) and later used for structure generation conditioned on crystal composition (Jiao et al., 2023; 2024; Miller et al., 2024; Levy et al., 2025). The use of generative models has spanned multiple modeling methods, including diffusion models with equivariant models (Jiao et al., 2023; 2024), symmetry-aware diffusion models (Levy et al., 2025), flow-matching models on specialized manifolds (Miller et al., 2024). Recent work have also applied large-language models to crystal generation and structure prediction with more varied success compared to other methods (Antunes et al., 2024; Ding et al., 2025).

H ADDITIONAL CHEMICAL ANALYSIS EXAMPLES

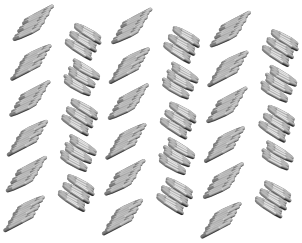


Figure 16: OXTAL learns small local neighborhoods from S^4 crops and generalizes to infer large and periodic structures. Example of ANTCEN with over 2400 tokens ($RMSD_{15}$ = 1.9Å)

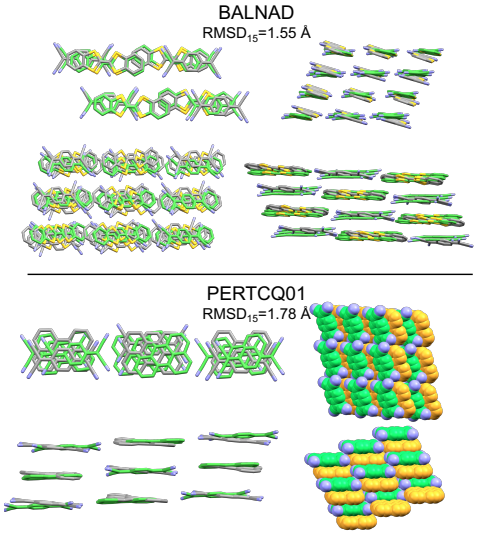


Figure 17: Examples of OXTAL generated co-crystal structures (green) compared against experimental structures (gray).

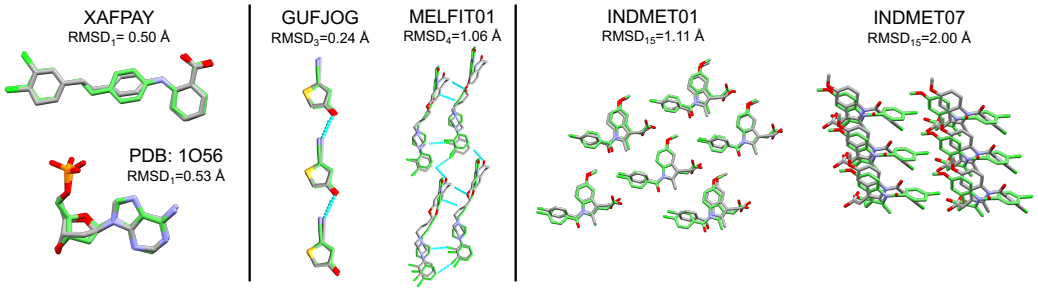


Figure 18: Examples of OXTAL generated structures (green) compared against experimental structures (gray). (A) flexible conformers, (b) intermolecular interactions, (c) polymorphs of the same molecule.

H.1 DIVERSITY OF GENERATED SAMPLES

In Figure 19, we provide a qualitative visualization of some example packings generated by OXTAL. We see that although several samples lie in the same orientation, some generated samples also present a more complex herringbone packing. This is evident in the XATJOT co-crystal, which does not collapse the two different molecular components into the same orientation. These results suggest that OXTAL may prefer planar packings, which are more prevalent in the training dataset. However, this may be mitigated with additional tuning on noise scheduler parameters.

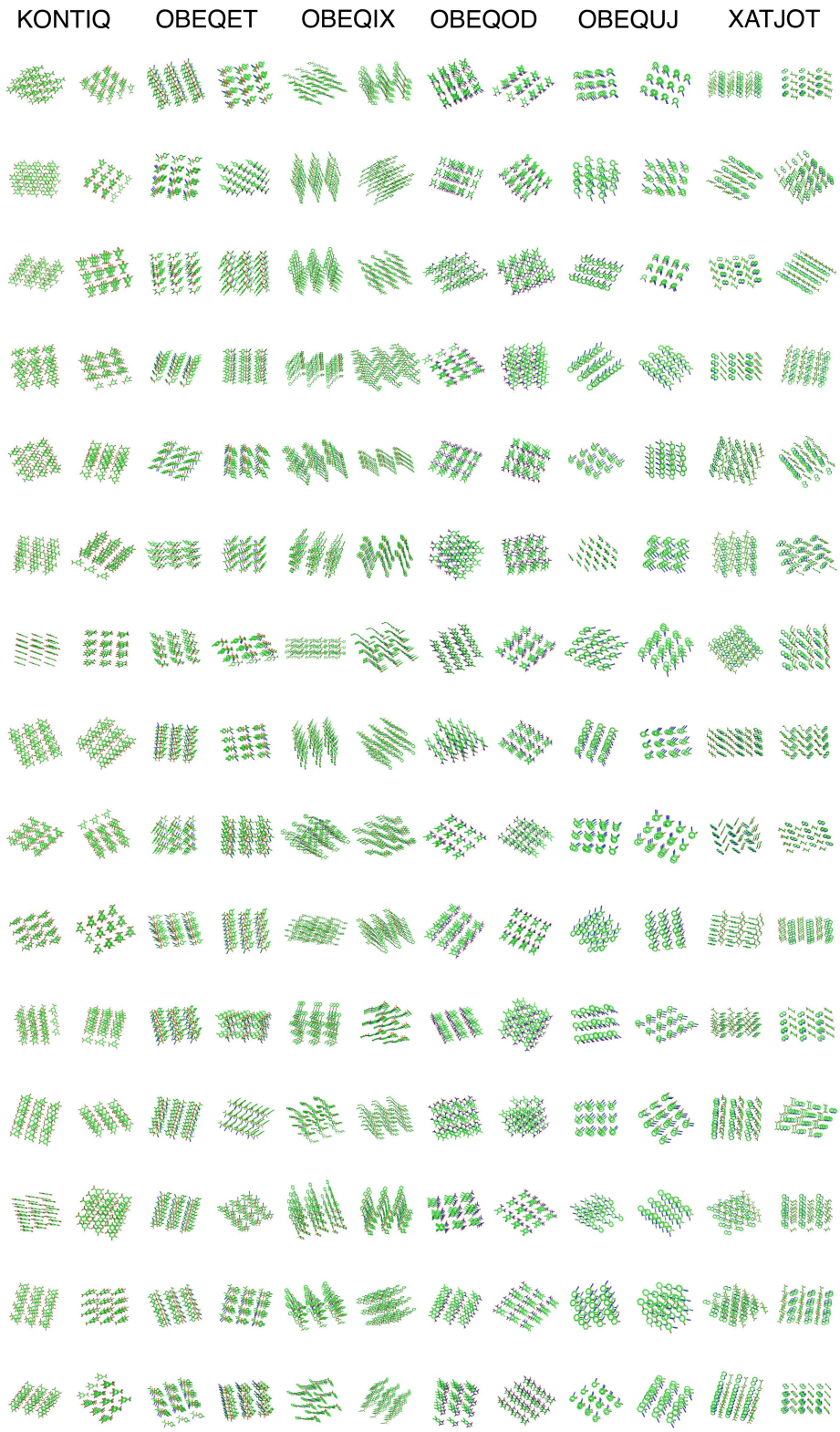


Figure 19: Example set of structures generated by OXTAL for various crystals (labeled by CSD ID).

I DETAILED DFT BASELINES FROM CSD BLIND TESTS

Table 9: Results per crystal of submitted methods for CSD blind test 5.

Model	n_S	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C?$	$\text{Rec}_S \uparrow$	$\text{Rec}_C?$	$\widetilde{\text{Sol}}_C?$
			KONTIQ				
Group Boerrigter	5002	0	0.896	Y	0.907	Y	Y
Group Day	2350	0	0.985	Y	0.987	Y	Y
Group Desiraju	244	0	0.951	Y	0.951	Y	Y
Group Hofmann	103	0	0.981	Y	0.981	Y	Y
Group Kendrick	1886	0	0.993	Y	0.994	Y	Y
Group Maleev	26	0	0.769	Y	0.769	Y	Y
Group Orendt	594	0	0.921	Y	0.936	Y	Y
Group Price	449	0.002	0.978	Y	0.980	Y	Y
			OBEQET				
Group Ammon	5	0	0	N	0.600	Y	N
Group Boerrigter	5001	0	0.005	Y	0.448	Y	Y
Group Day	438	0	0	N	0.893	Y	N
Group Desiraju	156	0	0.038	Y	1.000	Y	Y
Group Hofmann	103	0	0	N	0.379	Y	N
Group Jose	11	0.364	0	N	0	N	N
Group Kendrick	79	0	0.127	Y	0.911	Y	Y
Group Maleev	10	0	0	N	0.800	Y	N
Group Orendt	165	0	0.012	Y	0.897	Y	Y
Group Price	259	0	0.008	Y	0.525	Y	Y
			OBEQIX				
Group Ammon	11	0	0	N	0	N	N
Group Boerrigter	5005	0	0	N	0.014	Y	N
Group Hofmann	103	0.971	0	N	0	N	N
Group Kendrick	117	0	0.205	Y	0.248	Y	Y
Group Maleev	8	0	0	N	0	N	N
Group Orendt	156	0	0.006	Y	0.006	Y	Y
Group Price	103	0	0.039	Y	0.243	Y	Y
			OBEQOD				
Group Ammon	6	0	0.167	Y	1.000	Y	N
Group Boerrigter	4894	0	0.054	Y	0.999	Y	Y
Group Day	164	0	0.220	Y	1.000	Y	Y
Group Desiraju	202	0	0.302	Y	1.000	Y	Y
Group Hofmann	103	0	0	N	1.000	Y	N
Group Jose	15	0.533	0.133	Y	0.267	Y	N
Group Kendrick	150	0	0.200	Y	1.000	Y	Y
Group Maleev	24	0	0.083	Y	1.000	Y	N
Group Orendt	272	0	0.051	Y	0.971	Y	Y
Group Price	139	0	0.122	Y	1.000	Y	Y
			OBEQUJ				
Group Ammon	20	0	0.850	Y	1.000	Y	Y
Group Boerrigter	8456	0	0.113	Y	0.994	Y	Y
Group Day	549	0	0.222	Y	1.000	Y	Y
Group Desiraju	202	0	0.272	Y	1.000	Y	Y
Group Hofmann	103	0	0	N	1.000	Y	N
Group Jose	16	0.188	0.250	Y	0.438	Y	N
Group Kendrick	90	0	0.500	Y	1.000	Y	Y
Group Maleev	18	0	0	N	0.944	Y	N
Group Misquitta	103	0	0.214	Y	1.000	Y	Y
Group Nikylov	13	0	0	N	1.000	Y	N
Group Orendt	292	0	0.209	Y	1.000	Y	Y
Group Price	150	0	0.367	Y	1.000	Y	Y
			XATJOT				
Group Boerrigter	383	0	0.355	Y	1.000	Y	Y
Group Desiraju	202	0	0.262	Y	1.000	Y	Y
Group Kendrick	350	0	0.326	Y	1.000	Y	Y
Group Maleev	23	0	0.217	Y	1.000	Y	Y
Group Orendt	279	0	0.140	Y	1.000	Y	Y
Group Price	164	0	0.079	Y	1.000	Y	Y

Table 10: Results per crystal of submitted methods for CSD blind test 6.

Model	$n_{\text{submitted}}$	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C?$	$\text{Rec}_S \uparrow$	$\text{Rec}_C?$	$\widetilde{\text{Sol}}_C?$
			NACJAF				
Group Singh	100	0	0.010	Y	1.000	Y	N
Group Cole	100	0	0.050	Y	1.000	Y	Y
Group Day	200	0	0.085	Y	1.000	Y	Y
Group Dzyabchenko	100	0	0.170	Y	1.000	Y	Y
Group vanEijck	100	0	0.070	Y	1.000	Y	Y
Group FustiMolnar	198	0.005	0.056	Y	0.939	Y	Y
Group Cuppen	200	0	0.080	Y	1.000	Y	Y
Group Facelli	177	0	0.006	Y	0.977	Y	N
Group Obata	200	0	0.060	Y	1.000	Y	Y
Group Hofmann	100	0	0	N	1.000	Y	N
Group Ma	200	0	0	N	1.000	Y	N
Group Marom	200	0.045	0.030	Y	0.715	Y	Y
Group Mohamed	100	0	0.570	Y	1.000	Y	Y
Group Leusen	100	0	0.070	Y	1.000	Y	Y
Group Pantelides	100	0	0.040	Y	1.000	Y	Y
Group al	33	0	0	N	1.000	Y	N
Group Podeszwa	99	0	0.091	Y	1.000	Y	Y
Group Price	200	0	0.045	Y	1.000	Y	Y
Group Price	100	0	0.030	Y	1.000	Y	Y
Group Szalewicz	54	0	0.130	Y	1.000	Y	Y
Group Zhu	80	0	0.075	Y	0.988	Y	Y
Group Hofmann	31	0	0	N	1.000	Y	N
Group Grimme	119	0	0.042	Y	0.992	Y	Y
Group Tkatchenko	120	0	0.067	Y	0.992	Y	Y
			XAFPAY				
Group Singh	100	0	0	N	0	N	N
Group Cole	100	0	0.050	Y	0.090	Y	Y
Group Day	200	0	0.040	Y	0.035	Y	Y
Group vanEijck	100	0	0.040	Y	0.060	Y	Y
Group FustiMolnar	149	0	0.114	Y	0.094	Y	Y
Group Cuppen	200	0	0	N	0	N	N
Group Facelli	100	0	0	N	0	N	N
Group Obata	200	0	0.120	Y	0.135	Y	Y
Group Hofmann	100	0	0	N	0	N	N
Group Mohamed	100	0	0.090	Y	0.350	Y	Y
Group Leusen	200	0	0.170	Y	0.270	Y	Y
Group Pantelides	100	0	0.100	Y	0.120	Y	Y
Group Price	200	0	0.120	Y	0.165	Y	Y
Group Zhu	60	0	0.083	Y	0.133	Y	Y
Group Hofmann	18	0	0	N	0	N	N
Group Grimme	125	0	0.400	Y	0.304	Y	Y
Group Tkatchenko	50	0	0.280	Y	0.220	Y	Y
			XAFQAZ01				
Group Hofmann	100	0	0	N	1.000	Y	N
Group Tkatchenko	20	0	0.100	Y	1.000	Y	Y
			XAFQIH				
Group Singh	100	0	0	N	0	N	N
Group Cole	100	0	0.010	Y	0.010	Y	Y
Group Day	200	0	0.015	Y	0	N	Y
Group Dzyabchenko	71	0	0.014	Y	0	N	Y
Group vanEijck	100	0	0	N	0	N	N
Group FustiMolnar	138	0	0.312	Y	0.246	Y	Y
Group Hofmann	100	0	0	N	0	N	N
Group Mohamed	100	0	0	N	0	N	N
Group Leusen	200	0	0.405	Y	0.490	Y	Y
Group Pantelides	100	0	0.020	Y	0	N	Y
Group Price	200	0	0.035	Y	0.040	Y	Y
Group Zhu	30	0	0	N	0	N	N
Group Hofmann	15	0	0	N	0	N	N
Group Grimme	11	0	0	N	0.182	Y	N
			XAFQON				
Group Day	200	0.980	1.000	Y	0.020	Y	Y
Group vanEijck	100	0	0.980	Y	0.980	Y	Y
Group FustiMolnar	198	0	0.859	Y	0.005	Y	Y
Group Facelli	100	0.280	0.990	Y	0.720	Y	Y
Group Hofmann	100	0	0.990	Y	1.000	Y	Y
Group Leusen	100	0.310	1.000	Y	0.690	Y	Y
Group Price	200	0.805	0.990	Y	0.195	Y	Y
Group Zhu	50	0.140	1.000	Y	0.860	Y	Y
Group Hofmann	15	0.200	1.000	Y	0.800	Y	Y
Group Grimme	119	0.941	0.966	Y	0.059	Y	Y
Group Szalewicz	100	0.800	1.000	Y	0.200	Y	Y
Group Tkatchenko	50	0.600	1.000	Y	0.400	Y	Y

Table 11: Results per crystal of submitted methods for CSD blind test 7.

Model	n_S	$\text{Col}_S \downarrow$	$\text{Lat}_S \uparrow$	$\text{Lat}_C \uparrow$	$\text{Rec}_S \uparrow$	$\text{Rec}_C \uparrow$	$\widetilde{\text{Sol}}_C \uparrow$
			FASMEV				
Group XtalPi	1510	0	0.223	Y	0.997	Y	Y
Group Roza	319	0	0.245	Y	0.997	Y	Y
Group DKhakimov	80	0	0.537	Y	1.000	Y	Y
Group Isayev	1510	0	0.476	Y	1.000	Y	Y
Group SMohamed	1510	0	0.054	Y	0.984	Y	Y
Group Pantelides	1510	0	0.072	Y	0.961	Y	Y
Group MNeumann	1504	0	0.418	Y	0.999	Y	Y
Group Goto	1510	0	0.079	Y	0.738	Y	Y
Group CJPickard	1510	0.001	0.154	Y	0.994	Y	Y
Group SLPrice	1265	0	0.043	Y	0.999	Y	Y
Group CShang	1500	0	0.159	Y	0.988	Y	Y
Group QZhu	209	0	0.096	Y	0.947	Y	Y
Group DBoese	1510	0	0.060	Y	0.836	Y	Y
Group GDay	1510	0	0.143	Y	0.981	Y	Y
Group BEijck	1510	0	0.052	Y	0.668	Y	Y
			JEKVII				
Group XtalPi	1500	0.001	0.030	Y	0.020	Y	Y
Group SMohamed	203	0	0	N	0	N	N
Group Pantelides	1499	0.006	0.001	Y	0	N	N
Group MNeumann	1500	0	0.134	Y	0.066	Y	Y
Group SLPrice	1500	0	0	N	0	N	N
Group CShang	1500	0	0.085	Y	0.014	Y	Y
Group QZhu	1495	0	0	N	0	N	N
Group DBoese	1500	0	0	N	0	N	N
Group GDay	1500	0	0	N	0	N	N
Group BEijck	1500	0	0	N	0	N	N
			MIVZEA				
Group XtalPi	1600	0	0.042	Y	0.186	Y	Y
Group DKhakimov	281	0	0	N	1.000	Y	N
Group SMohamed	1600	0	0.002	Y	1.000	Y	Y
Group MNeumann	1600	0	0.042	Y	0.207	Y	Y
Group Goto	1600	0	0.002	Y	0.541	Y	Y
Group SLPrice	1600	0	0	N	0.269	Y	N
Group QZhu	1600	0	0.001	Y	0.070	Y	Y
Group GDay	1600	0	0.007	Y	0.379	Y	Y
Group BEijck	1600	0	0.007	Y	0.239	Y	Y
			MIVZIE				
Group XtalPi	1600	0	0.024	Y	0.107	Y	Y
Group DKhakimov	281	0	0.007	Y	0	N	N
Group SMohamed	1600	0	0	N	0	N	N
Group MNeumann	1600	0	0.015	Y	0.111	Y	Y
Group Goto	1600	0	0.006	Y	0.001	Y	Y
Group SLPrice	1600	0	0.001	Y	0	N	N
Group QZhu	1600	0	0.003	Y	0.083	Y	Y
Group GDay	1600	0	0.016	Y	0.099	Y	Y
Group BEijck	1600	0	0.004	Y	0.041	Y	Y
			OJIGOG01				
Group XtalPi	1500	1.000	0.141	Y	0	N	N
Group MNeumann	1500	0.947	0.127	Y	0.008	Y	N
Group SLPrice	1500	1.000	0.093	Y	0	N	N
Group CShang	1500	0.993	0.359	Y	0.005	Y	Y
Group KSzalewicz	1500	1.000	0	N	0	N	N
Group BEijck	1500	0.076	0.002	Y	0	N	Y
			XIFZOF01				
Group XtalPi	1500	0	0.061	Y	0.038	Y	Y
Group Isayev	1500	0	0.027	Y	0.005	Y	Y
Group Matsui	1500	0	0.005	Y	0.005	Y	Y
Group MNeumann	1500	0	0.015	Y	0.007	Y	Y
Group Goto	1500	0	0.011	Y	0.003	Y	Y
Group SLPrice	1500	0	0.005	Y	0	N	Y
Group CShang	1500	0	0.061	Y	0.007	Y	Y
Group QZhu	1500	0	0	N	0	N	N
Group BEijck	1500	0	0.011	Y	0.006	Y	Y
			ZEGWAN				
Group XtalPi	1500	0	0.029	Y	0.797	Y	Y
Group DKhakimov	56	0	0	N	0	N	N
Group Pantelides	1500	0	0.011	Y	0.429	Y	Y
Group MNeumann	1500	0	0.052	Y	0.729	Y	Y
Group Goto	1500	0	0.010	Y	0.553	Y	Y
Group SLPrice	1500	0	0.013	Y	0.363	Y	Y
Group CShang	1500	0	0.009	Y	0.545	Y	Y
Group QZhu	1453	0	0.012	Y	0.551	Y	Y
Group GDay	1500	0	0.013	Y	0.623	Y	Y
Group BEijck	1500	0	0.015	Y	0.338	Y	Y
			ZEHFUR				
Group XtalPi	1500	0	0.055	Y	0.513	Y	Y
Group Isayev	1500	0	0.063	Y	0.405	Y	Y
Group SMohamed	1500	0	0.001	Y	0	N	N
Group Pantelides	1500	0	0.029	Y	0.330	Y	Y
Group MNeumann	1500	0	0.123	Y	0.636	Y	Y
Group Goto	1500	0	0.006	Y	0.137	Y	Y
Group SLPrice	1500	0	0.017	Y	0.244	Y	Y
Group CShang	1500	0	0.069	Y	0.678	Y	Y
Group QZhu	1500	0	0.001	Y	0.045	Y	N
Group DBoese	1500	0	0.029	Y	0.266	Y	Y
Group GDay	1500	0.005	0.055	Y	0.351	Y	Y
Group BEijck	1500	0	0.008	Y	0.029	Y	Y

Table 12: Summary of computational resources used by some of the participants as reported in the 5th CSP Blind Test. CPU hours are approximately normalized to 3.0 GHz. OXTAL times are reported as elapsed wall time in hours on 1 L40s GPU and 6 CPUs. Inference time for OXTAL is negligible compared to traditional computation chemistry methods.

Group	XVI	XVII	XVIII	XIX	XX	XXI	Total
Boerrigter	90	100	350	650	2,105	600	3,800
Day, Cruz-Cabeza	110	1,941	21,051	6,097	54,090	22,197	91,400
Desiraju, Thakur, Tiwari, Pal	114	2,303	324	114		1,431	4,600
Hofmann	2	7	12	694	670	187	1,600
Neumann, Leusen, Kendrick, van de Streek							115,000
Price et al.	200	5,000	14,000	3,000	120,000	52,800	195,000
Van Eijck	27						9,500
Della Valle, Venuti							3,200
Maleev, Zhitkov							7,500
Misquitta, Pickard & Needs							162,000
Scheraga, Arnautova	150	150	610	720			1,300
OXtal	0.024	0.011	0.012	0.013	0.029	0.011	0.100

Table 13: Summary of the computational resources used by each submission in terms of raw CPU hours as reported in the 6th CSP Blind Test. OXTAL times are reported as elapsed wall time in hours on 1 L40s GPU and 6 CPUs. Inference time for OXTAL is negligible compared to traditional computation chemistry methods.

Group	XXII	XXIII	XXIV	XXV	XXVI	Total
Chadha & Singh	350	450			600	1,400
Cole et al.	6	538		46	246	836
Day et al.	12,714	394,948	15,241	121,701	179,897	724,501
Dzyabchenko	144	3,648	3,360			7,152
van Eijck	130	2,810	1,400	8,060	7,630	20,030
Elking & Fusti-Molnar	418,540	242,000	235,400	135,000	190,000	1,220,940
van den Ende, Cuppen et al.	9,741	7,777		6,388	23,906	
Facelli et al.	268,012	38,500	11,500	39,000		357,012
Obata & Goto	19,200	346,000		325,000		690,200
Hofmann & Kuleshova	10	630	623	202	255	1,720
Lv, Wang, Ma	325,000					325,000
Marom et al.	30,000,000					30,000,000
Mohamed	26	106		81	61	274
Neumann, Kendrick, Leusen	32,160	146,120	103,700	84,680	356,844	723,504
Pantelides, Adjiman et al.	333	87,000		37,535	272,500	397,368
Pickard et al.	380,000					380,000
Podeszwa et al.	72,220					72,220
Price et al.	26,000	84,000	63,000	169,000	327,000	669,000
Szalewicz et al.	66,000					66,000
Tuckerman, Szalewicz et al.	81,000					81,000
Zhu, Oganov, Masunov	4,000	275,000	279,800	30,000	180,000	768,800
Boese (Hofmann)	80,000	80,000	80,000	80,000	80,000	400,000
Brandenburg & Grimme (Price)	13,665	8,661	3,509	34,824	10,135	70,794
Szalewicz et al. (Price)			15,000			15,000
Tkatchenko et al. (Price)	100,000	2,100,000	500,000	500,000		3,200,000
OXtal	0.012	0.019	0.011	0.030	0.042	0.114

Table 14: CPU core hours per target molecule for each prediction method as reported in the 7th CSP Blind Test. OXTAL times are reported as elapsed wall time in hours on 1 GPU and 6 CPUs. Inference time for OXTAL is negligible compared to traditional computation chemistry methods.

Group	XXVII	XXVIII	XXIX	XXX	XXXI	XXXII	XXXIII	Total
1			652,495		840,000	1,597,000	412,000	3,501,495
3			1,600,000		1,500,000	3,600,000		6,700,000
5	768,766		33,000	2,900,000	510,563	846,698	228,957	5,287,984
6	8,120	1,350	1,310	9,800	1,470	2,900	4,980	29,930
8	3,200	10			4,000		1,840	9,050
10	772,500	1,242,500	1,146,588	644,927	381,672	644,927	612,500	5,445,614
11			643,882					643,882
12			20,000	80,000	20,000			120,000
13			350	1,500			500	2,350
16	1,700,000		2,128,000		630,000			4,458,000
17	95,819							95,819
18			1,050	36,864	632	1,561		40,107
19	30,000		40,000	1,250,000	140,000	400,000	60,000	1,920,000
20	1,022,976	283,538	755,712	1,769,472	1,028,064	3,935,232	728,064	9,523,058
21	333,586		92,890	580,436	1,889,649		477,210	3,373,771
22	20,000	2,000	15,000	180,000	20,000	25,000	25,000	287,000
23			10,000					10,000
24	450,290	89,666	76,541	100,000	49,177	244,520	123,427	1,133,621
25	55,150	29,691	4,784		6,476	76,161	34,648	206,910
26					28,332			28,332
27	1,280,566	60,457	242,424	213,722		1,663,940	150,650	3,611,759
28	1,600		1,500	7,680	1,500	1,500	1,500	15,280
OXtal	0.060	0.026	0.011	0.056	0.015	0.050	0.017	0.235