ATOMSTEER: TEST-TIME CONSTRAINT STEERING FOR ATOMISTIC GENERATIVE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion and flow matching models offer scalable approaches to generating molecular conformers, protein-ligand complexes, and crystalline materials. However, satisfying structural geometry, accurate stereochemistry, and crystallographic symmetry remains challenging. Satisfying these requirements calls for corrections that also preserve the geometry already captured by the pretrained model. In this work, we propose AtomSteer, a test-time inference framework for steering pretrained atomistic generative models with structural constraints. During sampling, AtomSteer corrects intermediate structure predictions with bounded local projections, graph-rigid stereochemical transformations, and projections over periodic symmetries, accepting updates that improve constraint satisfaction while limiting changes to the rest of the structure. We demonstrate AtomSteer across three pretrained backbones, ET-Flow, Boltz-2, and FlowMM, spanning molecular conformer generation, protein-ligand complex structure prediction, and de novo crystal generation and crystal structure prediction. AtomSteer improves molecular validity and *E*/*Z* stereochemical accuracy while retaining conformer ensemble fidelity. For protein-ligand complexes, AtomSteer reduces ligand *E*/*Z* errors relative to experimental reference structures while achieving competitive docking success and PoseBusters validity. AtomSteer improves space-group distribution coverage in de novo crystal generation and improves target space-group recovery in crystal structure prediction. Our method provides a unified framework for adding chemical and crystallographic control to existing atomistic generative models without retraining or architectural changes.

1 INTRODUCTION

Atomistic generative models now sample molecular conformers, protein–ligand complexes, and periodic crystals with broad distributional fidelity (Hassan et al., 2024; Passaro et al., 2025; Miller et al., 2024), yet useful samples must also satisfy structural information known before generation. Molecular graphs specify local geometry and stereochemistry, protein–ligand poses must remain chemically valid, and crystalline materials obey space-group symmetries that strongly influence their properties (Levy et al., 2025; Kazeev et al., 2025). These requirements remain imperfectly captured: recent co-folding models return wrong ligand chemistry or chirality for more than half of the tested ligands (Goddard & Croll, 2026; Ishitani & Moriwaki, 2025; Luthfi et al., 2026), and crystal generators, including FlowMM, over-produce low-symmetry structures such as $P1$ that are rare in nature (Levy et al., 2025).

Existing approaches build structural requirements into a representation, architecture, conditioning variable, training objective, or task-specific guidance or projection, which ties each to one model and requirement, and soft guidance only nudges samples toward lower violation (Section 2). What is missing is a common way to add heterogeneous structural requirements to a pre-trained atomistic generator. These requirements differ in what they measure and in how the structure should move: an *E*/*Z* error is measured on four atoms but corrected by rotating an entire substituent, a protein–ligand clash can be removed by moving the ligand rigidly, and crystal symmetry couples atomic orbits to the lattice. A single penalty or Euclidean projection treats these requirements alike, and task-specific methods are each built for one kind of requirement.

We introduce *AtomSteer*, a test-time framework that pairs the structural residual of each requirement on an atomistic graph with a structure-dependent projection that moves only the coordinates

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

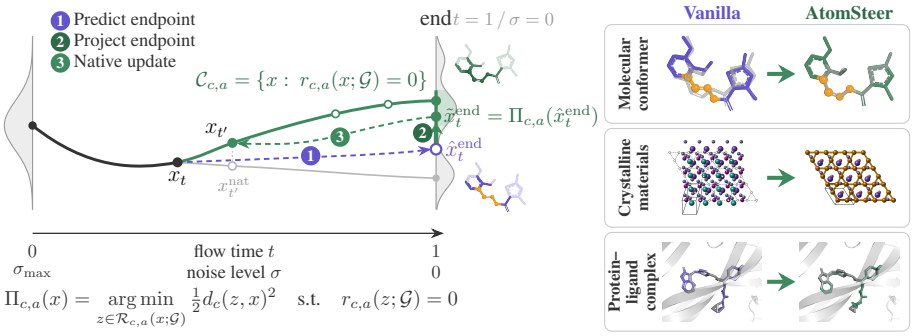


Figure 1: **AtomSteer applies structure-dependent projections during sampling.** (1) At a scheduled state x_t, the pre-trained generator provides the endpoint prediction $\hat{x}_t^{\text{end}}$ ($t = 1$ for flows, noise level $\sigma = 0$ for EDM). (2) AtomSteer projects it toward the structural constraint set $\mathcal{C}_{c,a}$ (Section 3.1), giving the projected endpoint $\tilde{x}_t^{\text{end}}$. (3) The native update from the projected endpoint gives $x_{t'}$ (green), departing from the unprojected native trajectory (gray). Fragments beside the end plane show a GEOM-Drugs *E*/*Z* double bond in its vanilla and AtomSteer conformers. Right: vanilla and AtomSteer samples for a molecular conformer, a crystal with a requested Archimedean lattice, and a protein–ligand complex; gray shows the ground truth.

the requirement concerns (Figure 1). AtomSteer projects the model's endpoint prediction at scheduled sampling steps, accepts only a bounded correction, and returns the projected endpoint through the pre-trained sampler's native update, without retraining or modifying the model. Three projection families cover this work: Chemical Gauss–Newton (ChemGN) for local geometry and stereochemistry, rigid-fragment and rigid-body projections for substituents and components, and crystal symmetry and motif projections for periodic structure. Herein, we test the same interface to steer FlowMM, ET-Flow, and Boltz-2, and projections from different families combine within one structure (Section 4.3).

Our contributions are:

1. **A general test-time framework for structural constraints on atomistic systems.** AtomSteer adds chemical and crystallographic constraints to pre-trained atomistic diffusion and flow models through bounded, structure-dependent endpoint projections without retraining, and recovers the native sampler's update when no correction is accepted.
2. **One formulation across atomistic systems.** Constraints are residuals on atomistic graphs, and three projection families cover local geometry, stereochemistry, component packing, periodic symmetry, and prescribed motifs, including symmetries inferred from endpoint predictions.
3. **Consistent gains on three representative models.** AtomSteer, applied on FlowMM, broadens de novo space-group coverage, raises composition-only MP-20 Joint@1 (structure match with the target space group) from 45.18% to 52.54%, and achieves a higher joint motif-constraint-and-prescreening rate than SCIGEN (Okabe et al., 2026) for all three Archimedean lattices. Applying on ET-Flow, AtomSteer improves conformer precision and PoseBusters validity on GEOM-Drugs, reduces median *E*/*Z* distribution error by 75.9%, and lowers GFN2-xTB ensemble-property errors by 10–36% relative to the better ET-Flow sampler for each property. Applying AtomSteer on Boltz-2, it achieves the highest PoseBusters validity on EZBench and raises joint valid-docking-and-stereochemistry success from 36.5% to 45.2% relative to Boltz-steering.

2 Related Work

Inference-time steering. Inference-time steering adapts a pre-trained diffusion or flow model by modifying sampling rather than parameters, and it is widely used to correct the ligands of co-folding models (Abramson et al., 2024; Wohlwend et al., 2025; Passaro et al., 2025; Team et al., 2024; 2026; Liu et al., 2024; Qiao et al., 2024; Team et al., 2025). Particle methods such as Feynman–Kac steering reweight whole trajectories against a reward (Singhal et al., 2025; Didi et al., 2026; Shen et al., 2025). Coordinate guidance modifies structures within single trajectories through differentiable potentials, restraints, or terminal optimization (Ishitani & Moriwaki, 2025;

Chen et al., 2026; Hori et al., 2026; Lai et al., 2026; He et al., 2023), and hybrid methods combine both (Ren et al., 2026; Wu et al., 2024; Scarpellini et al., 2026). Our Boltz-steering baseline applies seven flat-bottom potentials for chirality, stereochemistry, internal geometry, and clashes as gradient guidance on the denoised estimate, together with Feynman–Kac resampling (Wohlwend et al., 2025; Passaro et al., 2025). Hard-projection methods act on noisy samples (Christopher et al., 2024) or on denoised intermediates and predicted endpoints (Utkarsh et al., 2025; Christopher et al., 2026). Closest to our work on protein–ligand complexes, ProteinGS projects onto physical validity with a Gauss–Seidel solver once, on the finished structure, with the denoiser fine-tuned through the projection (Chen et al., 2026). AtomSteer pairs the structural residual of each atomistic graph constraint with a structure-dependent projection whose corrections act through local Cartesian, rigid-fragment, rigid-body, or periodic coordinates. It projects endpoint predictions at intermediate sampling steps and returns accepted corrections through the native update without fine-tuning, across molecular conformers, protein–ligand complexes, and periodic crystals.

Symmetry and motifs in crystal generation. Symmetry-aware crystal generators build space-group or chemical constraints into their representation, conditioning, or training: DiffCSP++ constrains diffusion to the lattice and Wyckoff positions of a supplied space group (Jiao et al., 2024), SymmCD generates asymmetric units with their symmetry transformations (Levy et al., 2025), and CrysVCD filters language-model compositions for charge balance before diffusion (Cheng et al., 2025). SCIGEN imposes requested motifs at inference by masking a diffusion model's denoising path (Okabe et al., 2026). AtomSteer instead corrects FlowMM's endpoint prediction at test time, inferring a compatible symmetry without the target structure, a supplied space group, or an energy model, and imposing requested motifs through the same projection interface.

Molecular geometry and stereochemistry. Graph-, fragment-, and torsion-based generators encode geometry during training. GeoMol predicts local structures and torsion angles, Torsional Diffusion samples rotatable torsions, and SigmaDock diffuses rigid-fragment poses in $SE(3)$ (Ganea et al., 2021; Jing et al., 2022; Prat et al., 2026). Stereochemistry can also be encoded through targeted objectives, graph augmentation, or training data (Williams & Inala, 2024; Nikitin et al., 2026; Lin et al., 2025; Vilya Research et al., 2026). AtomSteer instead projects graph-defined requirements onto the endpoint predictions of a pre-trained Cartesian generator during sampling.

3 AtomSteer: Test-Time Constraint Projection

Pre-trained atomistic generative models learn useful structural distributions, but their samples are not guaranteed to satisfy additional chemical or crystallographic requirements. AtomSteer adds such structural constraints to a pre-trained atomistic sampler at test time (Figure 1). Each constraint pairs a structural residual, which defines what must be satisfied, with a structure-dependent projection, which defines which coordinates move together to satisfy it. AtomSteer corrects the model's *endpoint prediction* $\hat{x}_t^{\mathrm{end}}$ of the final structure rather than the noisy state x_t, and the native sampler turns the projected endpoint $\tilde{x}_t^{\mathrm{end}}$ into the next state $x_{t'}$, so later sampling steps respond to the correction (Figure 1).

3.1 Structural residuals and structure-dependent projections

Let $\mathcal{G} = (V, \mathcal{E}, \xi)$ be an atomistic graph, where its vertices V are atoms or crystal sites. Its relations $\mathcal{E}$ group the atoms that a constraint involves, such as a bonded pair, the four atoms around a double bond, the atoms of one molecule, or symmetry-related crystal sites. The labels ξ carry properties such as bond orders, stereochemical labels, and atomic species. The configuration x contains the generated variables. For a constraint c and a discrete realization a, such as one space-group setting, we define a violation residual and its constraint set by

$$r_{c,a}(x; \mathcal{G}), \qquad \mathcal{C}_{c,a} = \{x : r_{c,a}(x; \mathcal{G}) = 0\}. \tag{1}$$

For example, a steric residual measures how far two atoms fall below their minimum separation, and a periodic-symmetry residual measures the disagreement between an atom and its symmetry-related structure.

A structure-dependent projection reduces these violations by moving the appropriate coordinates together. With $\mathcal{R}_{c,a}(x; \mathcal{G})$ the structures reachable through these moves, it seeks the

Table 1: **Structural constraints grouped by projection family.** One requirement can combine families: *E*/*Z* pairs rigid-fragment basin correction with ChemGN local refinement, and protein–ligand exclusion can use ChemGN or rigid ligand motion. Appendices A, B, and C.2 define the residuals and details.

Projection family	Structural constraint examples	Coordinates corrected	Projection
Chemical Gauss–Newton (ChemGN)	Local chemical geometry and sterics: bond and 1–3 geometry, designated planarity, intramolecular clashes, protein–ligand exclusion, and local *E*/*Z* refinement	Selected molecular or ligand coordinates	Cartesian ChemGN
	Local *R*/*S* handedness	Ligand coordinates with first-order bond protection	Bond-tangent ChemGN
Rigid-fragment and rigid-body projections	*E*/*Z* configuration	Connected substituent around the double bond	Rigid-fragment flip and refinement
	Component packing	Ligand or eligible protein component	Rigid-body $SE(3)$
Crystal symmetry and motif projections	Space-group symmetry	Fractional coordinates and lattice	Hierarchy-aware symmetry selection, orbit projection, and lattice-metric projection
	Lattice-metric relations	Lattice	Metric-template projection
	Requested Archimedean lattice	Selected species, fractional positions, and lattice entries	Motif projection

constraint-satisfying structure with minimum displacement,

$$\Pi_{c,a}(x) = \operatorname*{arg\,min}_{z \in \mathcal{R}_{c,a}(x;\mathcal{G})} \frac{1}{2} d_c(z,x)^2 \quad \text{subject to} \quad r_{c,a}(z;\mathcal{G}) = 0, \tag{2}$$

within the projection's movement budget. Bounded numerical corrections use the corresponding residual–displacement objective,

$$\min_{z \in \mathcal{R}_{c,a}(x;\mathcal{G})} \frac{1}{2} \|r_{c,a}(z;\mathcal{G})\|_{W_c}^2 + \frac{\lambda_c}{2} d_c(z,x)^2. \tag{3}$$

The distance d_c measures structural displacement, W_c weights the residuals, and λ_c controls the displacement penalty. Table 1 organizes the three projection families that realize these objectives by the coordinates each corrects. These projections instantiate one interface: each supplies a residual and a structure-dependent projection, and all share the same endpoint intervention, bounded acceptance, and native update. On ET-Flow and Boltz-2 the molecular projections use the same implementation.

Molecular stereochemistry and requested crystal motifs specify their targets directly. When a crystal's space group is not supplied, AtomSteer infers it from the model's own endpoint prediction through hierarchy-aware symmetry selection. This selection enumerates the space groups that the endpoint prediction already approximately satisfies and promotes along minimal-supergroup relations to the most symmetric feasible parent group. AtomSteer then projects the atomic positions onto symmetry orbits of that group in fractional coordinates, and the lattice metric onto the group's metric constraints, with atom types fixed (Appendix B.1.1).

3.2 Endpoint projection during pre-trained sampling

For the three pre-trained backbones we apply AtomSteer on, which span Cartesian flow matching, Riemannian flow matching, and EDM-style denoising diffusion, the endpoint predictions are

$$\hat{x}_t^{\text{end}} = \begin{cases} x_t + (1-t)v_\theta(x_t,t), & \text{ET-Flow}, \\ \text{projx}[x_t + (1-t)v_\theta(x_t,t)], & \text{FlowMM}, \\ D_\theta(x_\sigma,\sigma), & \text{Boltz-2}. \end{cases} \tag{4}$$

The flow models extrapolate the raw learned velocity v_θ to $t = 1$, with projx applying FlowMM's representation operations, including periodic wrapping. Boltz-2 follows EDM, in which σ is the noise level of unscaled noisy coordinates $x_\sigma = x_0 + \sigma\epsilon$ with $\epsilon \sim \mathcal{N}(0, I)$, and its denoiser D_θ maps x_σ to an estimate of the clean ($\sigma = 0$) coordinates x_0 (Karras et al., 2022; Passaro et al., 2025). Its step from t to t' below is one EDM update (Appendix C.2).

At scheduled steps, the projected endpoint replaces the endpoint prediction in the model's native update from x_t to $x_{t'}$. Each projection bounds its correction, because fully satisfying one constraint can distort nearby geometry, such as the bonds next to a removed steric clash. A correction that fails its acceptance checks is discarded, and those coordinates keep their native update in the original pretrained denoiser. Appendices B.1, B.2, and C.2 give each backbone's instantiation and schedule, and Appendix A defines the shared molecular projections.

4 RESULTS

We apply AtomSteer to three pre-trained models with a broad range of structural constraint types (Section 3.2). For each experiment, we ask whether the targeted constraint is better satisfied and whether a scientifically relevant outcome improves with it (protocols in Appendices B, C, and D).

4.1 ATOMSTEER IMPROVES CRYSTALLOGRAPHIC SYMMETRY IN FLOWMM

On FlowMM (Miller et al., 2024), the tasks progress from global symmetry in MP-20 (Horton et al., 2025) *de novo* generation, to target symmetry in crystal structure prediction (CSP) on MP-20, to requested local motifs (Okabe et al., 2026).

De novo crystal generation. We ask whether AtomSteer can move FlowMM's space-group distribution toward MP-20 without a supplied target group. Hierarchy-aware symmetry selection promotes the symmetry that FlowMM's endpoint prediction already approximates to the most symmetric feasible parent group and projects the atomic symmetry orbits and lattice metric onto it (Section 3.1).

AtomSteer moves the generated space-group distribution toward MP-20, especially with improvements for cells with more atoms. AtomSteer raises the symmetry success rate from 71.47% to 77.91%, broadens the generated space-group support, and lowers d_{SG} (Table 2; Figure 2a). The gain concentrates in unit cells with more atoms, where vanilla FlowMM most often loses symmetry (Figure 2b), and persists after full-cell UMA relaxation (Table 2; Appendix E). AtomSteer therefore recovers symmetry that the pre-trained trajectory already approaches, moving the generated space-group distribution toward MP-20 without a supplied target group.

Crystal structure prediction. For CSP, we ask whether AtomSteer can recover the target symmetry together with the crystal structure of a given composition. **Structure match alone can hide crystallographic errors.** Benchmarks usually score CSP by the StructureMatcher match rate (MR@K) (Jiao et al., 2023; Miller et al., 2024), so we also report target-space-group recovery (SG@K) and Joint@K, which requires one candidate to satisfy both. CrystalFlow (Luo et al., 2025)

Table 2: **(a) AtomSteer improves crystallographic symmetry in MP-20 *de novo* generation.** Metrics use 10,000 paired samples per method. Symmetry success rate is the fraction of all samples assigned a space-group number of at least three, d_{SG} is the Jensen–Shannon distance to the MP-20 space-group distribution, Unique SGs counts distinct detected space groups, Valid is the as-generated geometric validity rate, and $\Delta F_{\max}$ is the median paired change in maximum atomic force relative to FlowMM (not ranked). Relaxed columns use full-cell UMA relaxation (Wood et al., 2025) of the same endpoints. Best **bold**, second-best underlined. Details: Appendices B.1, D.1.1, and E.

	As generated					*After UMA relaxation*	
Method	Symmetry success rate (%) ↑	d_{SG} ↓	Unique SGs ↑	Valid (%) ↑	Paired $\Delta F_{\max}$ (eV/Å) ↓	Symmetry success rate (%) ↑	d_{SG} ↓
FlowMM	71.47	0.5541	89	**97.82**	–	64.19	0.5363
FlowMM + AtomSteer	**77.91**	**0.5083**	**105**	97.79	−0.0394	**71.50**	**0.5095**

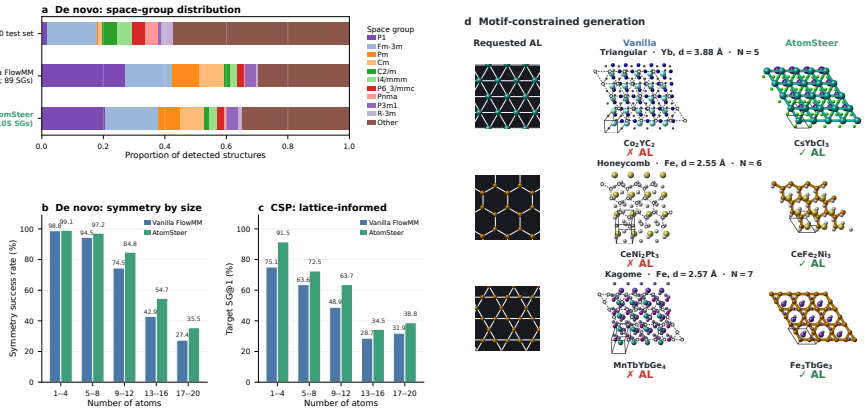


Figure 2: **AtomSteer improves crystallographic symmetry and realizes requested structures in FlowMM. a,** Detected space-group distributions of the MP-20 test set, vanilla FlowMM, and AtomSteer in *de novo* generation. **b,** Symmetry success rate of paired *de novo* samples by number of atoms in the generated unit cell. **c,** Target-space-group recovery (SG@1) for lattice-informed MP-20 CSP by number of atoms. **d,** Requested triangular, honeycomb, and Kagome Archimedean lattices (ALs) beside matched vanilla FlowMM and AtomSteer endpoints; dashed guides in the vanilla panels mark where AtomSteer realizes the AL. Metrics, denominators, and tolerances: Appendix D.1.

Table 3: **(b) Crystal structure prediction on MP-20** ($N = 9{,}046$)**.** Values (%) are structure match (MR@K), target-space-group recovery (SG@K), and both on the same sample (Joint@K) within K samples. † marks MR values reported in the original papers (Xie et al., 2022; Jiao et al., 2023; Luo et al., 2025; Höllmer et al., 2025; Seong et al., 2026; Miller et al., 2024); all other rows use the common evaluator. AtomSteer (lattice-informed) also receives a reference-derived lattice metric template and integer setting; all other rows receive composition only. Best values are bold and second-best underlined. Definitions and evaluation details: Appendix D.1.2.

	MR ↑		SG ↑		Joint ↑	
Method	@1	@20	@1	@20	@1	@20
CDVAE†	33.90	66.95	–	–	–	–
DiffCSP	52.21	77.77	47.56	75.41	41.50	65.49
CrystalFlow	62.15	<u>78.27</u>	3.06	15.03	1.64	12.77
OMatG	63.50	74.38	1.86	2.30	0.63	1.35
MCFlow-L†	<u>64.08</u>	76.08	–	–	–	–
FlowMM	62.45	75.29	51.66	71.17	45.18	62.44
FlowMM + AtomSteer (composition-only)	62.26	74.94	<u>59.83</u>	<u>75.81</u>	<u>52.54</u>	<u>66.32</u>
FlowMM + AtomSteer (lattice-informed)	**66.33**	**81.65**	**62.68**	**80.69**	**54.95**	**70.61**

and OMatG (Höllmer et al., 2025) reach MR@1 on par with FlowMM, yet attain SG@1 of only 3.06% and 1.86% (Table 3).

AtomSteer (composition-only) improves space-group and joint recovery while preserving match rate. With the symmetry inferred as in *de novo* generation, AtomSteer raises FlowMM's MP-20 Joint@1 from 45.18% to 52.54% while MR@1 is essentially unchanged, with the highest SG and Joint values among composition-only methods at $K = 1$ and 20 (Table 3). Per-size results are in Appendix E.

Lattice-system information further improves structure and joint recovery. Powder diffraction typically determines the crystal system before the atomic structure is solved (Park et al., 2017; Suzuki et al., 2020). AtomSteer (lattice-informed) receives only this discrete information, the reference lattice system and its integer conventional-cell setting, while FlowMM still generates the numerical cell parameters and atomic positions and hierarchy-aware symmetry selection infers the space group (Appendix B.1.1). It then attains the highest value in every column of Table 3 (SG@1 by size in Figure 2c), including 66.33% MR@1 and 54.95% Joint@1. In both information regimes, AtomSteer therefore adds crystallographic correctness to FlowMM without sacrificing structure match.

Table 4: **(c) Motif-constrained generation with AtomSteer on FlowMM against SCIGEN.** Each row uses all 1,000 planned conditions for its Archimedean lattice (AL). All metrics use our common evaluator. Joint requires AL constraint satisfaction and prescreening survival on the same sample, and AL detected after relaxation requires a valid, converged UMA endpoint. Best values within each AL and metric are bold and second-best underlined. Protocols: Appendices D.1.3 and B.1.

AL	Method	AL constraint satisfied (%) ↑	Prescreening survival (%) ↑	Joint (%) ↑	AL detected after relaxation (%) ↑
Kagome	SCIGEN	<u>98.3</u>	8.0	<u>8.0</u>	<u>10.3</u>
	FlowMM (vanilla)	0.1	**45.2**	0.1	0.3
	AtomSteer	**98.8**	<u>11.4</u>	**10.9**	**10.5**
Honeycomb	SCIGEN	<u>98.4</u>	7.4	<u>7.4</u>	**13.1**
	FlowMM (vanilla)	0.0	**47.3**	0.0	0.0
	AtomSteer	**99.2**	<u>8.5</u>	**7.7**	<u>9.4</u>
Triangular	SCIGEN	**97.5**	2.5	<u>2.5</u>	<u>31.5</u>
	FlowMM (vanilla)	0.1	**61.4**	0.1	2.5
	AtomSteer	<u>96.6</u>	<u>33.1</u>	**30.7**	**86.4**

Table 5: **AtomSteer improves conformer precision, validity, *E/Z* agreement, and GFN2-xTB ensemble properties on GEOM-Drugs. a,** Equal-molecule averages over the GEOM-Drugs test set with $2K$ generated conformers for K references. COV (0.75 Å threshold) and AMR (Å) are coverage and mean minimum RMSD over references (R) or generated conformers (P); PB-valid is the PoseBusters validity rate of the first $\min(2K, 32)$; and *E/Z* W_1 is the Wasserstein distance to the GEOM dihedral distributions, median over the 153 molecules with an evaluable stereogenic double bond. **b,** Median absolute errors of Boltzmann-weighted GFN2-xTB (Bannwarth et al., 2019) ensemble properties (mean energy E, dipole moment μ, HOMO–LUMO gap $\Delta\epsilon$, minimum energy $E_{\min}$) on a fixed random 100-molecule panel. ODE50 and SS50 are ET-Flow's released deterministic and stochastic samplers, and AtomSteer steers the same pre-trained ET-Flow checkpoint. Best **bold**, second best <u>underlined</u>. Definitions and protocols are in Appendices D.2 and B.2.

	a. Conformer generation						**b. GFN2-xTB ensemble properties**			
Method	COV-R ↑	COV-P ↑	AMR-R ↓	AMR-P ↓	PB-valid ↑	*E/Z* W_1 ↓ (°)	E ↓ (kcal/mol)	μ ↓ (D)	$\Delta\epsilon$ ↓ (kcal/mol)	$E_{\min}$ ↓ (kcal/mol)
ODE50	<u>79.92</u>	73.40	0.458	0.562	93.13	9.04	<u>0.1795</u>	0.2245	<u>0.3397</u>	0.0217
SS50	**79.93**	<u>73.52</u>	**0.456**	<u>0.559</u>	<u>93.17</u>	<u>8.94</u>	0.1812	<u>0.2193</u>	0.3413	<u>0.0197</u>
AtomSteer	79.01	**75.49**	<u>0.457</u>	**0.529**	**96.93**	**2.15**	**0.1610**	**0.1558**	**0.2354**	**0.0126**

Motif-constrained generation. Magnetic atoms on triangular, honeycomb, or Kagome lattices can host exotic states such as quantum spin liquids (Okabe et al., 2026). We therefore ask whether AtomSteer can impose on FlowMM one of these three primary Archimedean lattices (ALs), with a specified magnetic element on its vertices at a specified bond length, scoring stability with SCIGEN's released prescreening filter (Appendix D.1.3). AtomSteer projects FlowMM's endpoint prediction onto the requested AL by setting the vertex species, vertex positions, and in-plane lattice (Appendix B.1.2).

AtomSteer turns unconstrained FlowMM into an AL-constrained generator. Vanilla FlowMM inference almost never produces the requested AL, whereas AtomSteer realizes it in 96.6–99.2% of samples, on par with SCIGEN (Table 4; Figures 2d and 8). It also attains the highest joint rate for all three ALs, and the requested triangular AL is detected after relaxation in 86.4% of its samples, against 31.5% for SCIGEN. The same framework therefore extends from inferred global symmetry to explicitly requested local motifs.

4.2 AtomSteer improves molecular validity, stereochemistry, and conformer quality in ET-Flow

We ask whether enforcing the local geometry and stereochemistry also improves ET-Flow's conformer ensembles and their quantum-chemical properties. AtomSteer applies rigid-fragment *E/Z* projections, ChemGN, and a terminal projection to ET-Flow on the GEOM-Drugs test set (Axelrod & Gómez-Bombarelli, 2022) (Appendix B.2).

Table 6: **Protein–ligand steering results** on the PoseBusters benchmark (301 targets) and EZBench (1,103 targets). All entries are percentages except median *E*/*Z* error. All methods are inference-time steering approaches applied to the same pre-trained Boltz-2. Best **bold**, second-best underlined; ties share the mark. RMSD < 2 Å is docking success, and Success@1 and Success@5 require a PB-valid pose with RMSD < 2 Å (confidence top-1; any of five). Definitions: Appendix C.1.

	PoseBusters benchmark			EZBench					
Metric	Vanilla	Boltz-steering	**AtomSteer**	Vanilla	Native potentials	Boltz-steering	Ishitani & Moriwaki	ProteinGS	**AtomSteer**
PB-valid ↑	79.1	**99.3**	**99.3**	51.0	90.7	91.7	67.6	87.1	**93.4**
RMSD < 2 Å ↑	83.7	84.1	**84.7**	63.2	64.0	64.5	62.2	64.2	**64.7**
Success@1 ↑	66.1	83.4	**84.1**	35.7	59.1	59.7	45.7	56.8	**60.7**
Success@5 ↑	78.3	87.3	**88.7**	46.9	66.0	65.8	57.9	63.3	**67.5**
E/*Z* median error (°) ↓	6.12	5.26	**2.98**	6.27	6.44	6.13	11.13	6.73	**2.61**
All *E*/*Z* bonds $\leq 10°$ ↑	61.9	57.1	**76.2**	59.6	59.2	59.5	42.8	58.1	**79.1**
All *R*/*S* centers correct ↑	73.4	**100.0**	**100.0**	51.7	85.3	85.3	**85.7**	85.1	85.1
Success@1+stereo ↑	59.5	81.1	**82.7**	25.4	36.5	36.5	24.0	36.1	**45.2**

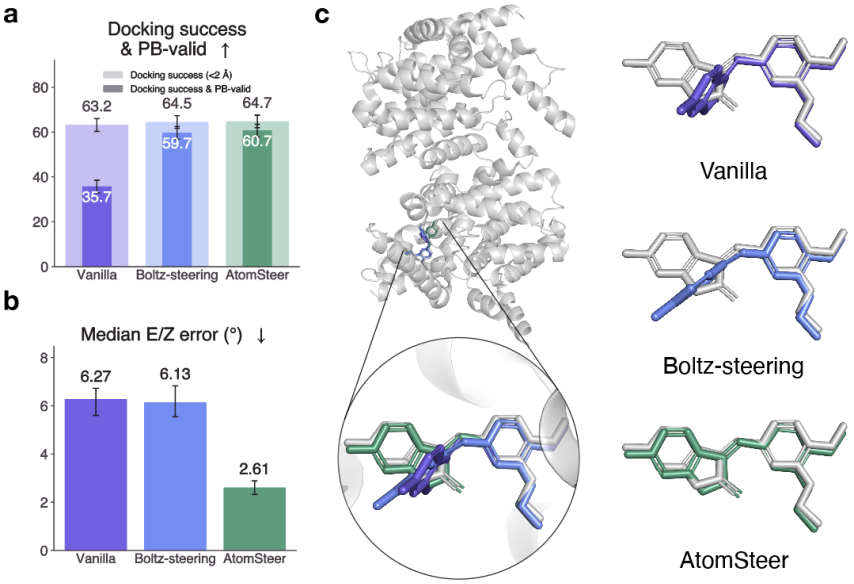


Figure 3: **AtomSteer corrects ligand stereochemistry on EZBench with improved docking success. (a)** Docking success (ligand RMSD < 2 Å after pocket alignment, light) and Docking success & PB-valid (Success@1, dark) on EZBench. **(b)** Median per-ligand *E*/*Z* dihedral error to the deposited ligand on EZBench. **(c)** Example complex (PDB `8ZQ2`): deposited ligand in gray, each method's pose overlaid, inset on the stereogenic bond.

AtomSteer improves stereochemistry, validity, and conformer precision while retaining ensemble coverage. Relative to the stronger ET-Flow sampler, SS50, AtomSteer lowers the median *E*/*Z* distribution error by 75.9%, raises PoseBusters validity to 96.93%, and improves COV-P and AMR-P, while COV-R stays at 79.01% against 79.93% (Table 5a; examples in Appendix Figures 9 and 10).

The geometry corrections also improve quantum-chemical ensemble properties. After relaxation with GFN2-xTB, AtomSteer attains the lowest median error on all four ensemble properties (as evaluated in the original ET-Flow paper), 10–36% below the better ET-Flow sampler for each property (Table 5b; Appendix D.2.3). These gains show that enforcing graph-specified structure can improve chemically relevant ensemble observables rather than merely pass geometric checks.

4.3 AtomSteer improves *E*/*Z* geometry and physical validity in Boltz-2

Co-folding models receive ligand stereochemistry in the input graph, yet their predicted complexes often invert or distort it. Existing steering targets physical validity and tetrahedral chirality, so we test whether AtomSteer can correct the ligand geometry and stereochemistry on Boltz-2 (Passaro et al.,

2025), including stereogenic double bond stereochemistry. AtomSteer enforces these graph-derived constraints with rigid-fragment *E*/*Z* projections, ChemGN, and rigid-body projections (Section 3.1).

We evaluate on the PoseBusters benchmark (Buttenschoen et al., 2024) and on EZBench, 1,103 X-ray protein–ligand complexes whose ligands each contain at least one stereogenic double bond, to investigate performance on *E*/*Z* isomerism in more depth, against unsteered Boltz-2 (Vanilla) and Boltz-steering. On EZBench we add native potentials alone, the stereochemical restraint guidance of Ishitani & Moriwaki (2025), and ProteinGS inference-only projections (Chen et al., 2026) on Boltz-2. One primary *E*/*Z* metric is the fraction of ligands whose every stereogenic double bond lies within 10° of the deposited dihedral. The discrete *E*/*Z* descriptor, which the PoseBusters double bond check compares, is saturated at 99.4–99.6% for most steered methods and does not measure how far the dihedral deviates from the experimental structure. Metric definitions and denominators are in Appendix C.1, and benchmark information and baseline settings in Appendices C.3 and C.2.

Existing methods repair validity and chirality, but not double bond geometry. On EZBench, every baseline substantially raises PoseBusters validity and *R*/*S* accuracy, yet none raises the within-10° fraction above its unsteered 59.6%, and Ishitani & Moriwaki, the most accurate on *R*/*S*, lowers it to 42.8% (Table 6).

AtomSteer corrects *E*/*Z* geometry. On EZBench, AtomSteer raises the within-10° fraction to 79.1%, $+19.7$ points over Boltz-steering, and more than halves the median error, from 6.13° to 2.61° (Table 6; Figure 3b). The same correction holds on the *E*/*Z* subset of the PoseBusters benchmark. In Figure 3c, these three methods assign the correct descriptor but predicted inaccurate dihedrals compared to the experimental geometry. Appendix Figures 4, 5, and 6 show further corrections, unchanged already-correct cases, and cases where AtomSteer displaces the ligand without a stereochemical gain.

AtomSteer attains the best validity and docking success, and the highest joint success. On both benchmarks, AtomSteer has the highest PoseBusters validity, docking success, Success@1 (confidence top-1), and Success@5 (by ligand RMSD) of all evaluated methods (Table 6). On EZBench, PoseBusters validity rises from 91.7% under Boltz-steering to 93.4%, Success@5 improves, and *R*/*S* accuracy is at parity (Table 6). Success@1+stereo, which requires the confidence top-1 pose to be PB-valid, docked below 2 Å, within 10° at every stereogenic double bond, and correct at every stereocenter, rises from 36.5% under Boltz-steering to 45.2%. Protein structure quality is largely indistinguishable across methods (Table 12).

The *E*/*Z* gain holds after the training cutoff. On the 177 EZBench complexes released after the Boltz-2 training cutoff, docking success falls for every method, whereas the *E*/*Z* margin over Boltz-steering stays at $+19.2$ points ($+19.7$ on the full set; Table 11), consistent with the correction enforcing graph-specified chemistry rather than relying on a seen complex. On the PoseBusters benchmark, whose entries all predate the cutoff, both steered methods reach 99.3% validity and 100% *R*/*S* accuracy, and AtomSteer raises the within-10° fraction from 57.1% to 76.2% on the 21 ligands with an evaluable double bond. AtomSteer therefore corrects the stereochemical failure it targets while preserving the pose and folded structure produced by the pre-trained model.

5 Conclusion

AtomSteer provides a common test-time interface for structural constraints in pre-trained atomistic generative models. Each requirement pairs a structural residual with a structure-dependent projection, and AtomSteer projects the model's endpoint prediction, accepts only a bounded correction, and returns it through the native denoising sampler, so adding a requirement means adding a constraint function, not retraining the model. With this single interface, AtomSteer improves crystallographic symmetry and motif realization, conformer stereochemistry and quantum-chemical ensemble properties, and protein–ligand *E*/*Z* geometry, validity, and joint success over their pre-trained models. AtomSteer currently requires a generative model that exposes, or permits a useful estimate of, its predicted endpoint during generation, and each requirement must be expressible as a structural residual with a corresponding projection. Extending the framework to broader diffusion and flow parameterizations, and to learned projections for less structured requirements, are natural next steps. AtomSteer can also accommodate further constraints, from coordination environments and oxidation states to experimentally derived structural information, enabling increasingly controllable generation across chemistry and materials design.

AI USE STATEMENT

In this work, we used generative AI tools for running experiments and editing code, drafting parts of the paper, and plotting results. We have not used generative AI tools for helping to develop theoretical models and conceptual frameworks, or to implement methods, or to propose and refine hypotheses, or for the design of or feedback on research methodology and experiments, or interpreting results, or cleaning and reformatting datasets, or qualitative or thematic analysis. Generation of synthetic data sets, translation, formulation of mathematical claims, and providing ingredients for or writing proofs are not applicable to this work. We have reviewed all AI-assisted work. All AI-generated code was reviewed line by line by at least one author. The implementation of our proposed method was additionally checked against its mathematical description in the paper. All figures were regenerated from raw experiment logs, and reported numbers were cross-checked against those logs. AI-drafted text was edited by the authors; every factual claim was checked, and every citation was verified against the original source. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We have put a lot of effort into ensuring that our results are reproducible. The AtomSteer framework, its constraint residuals, and its projection families are defined in Section 3 and Appendix A. Their integration with each pre-trained backbone, including endpoint prediction, native update, projection order, and schedules, is summarized in Table 7 and Algorithm 1 and specified in Appendix B.1 for FlowMM, Appendix B.2 for ET-Flow, and Appendix C.2 for Boltz-2. All experiments use frozen pre-trained backbone checkpoints and the released checkpoints of the baselines, with the versions, checkpoints, and sampling settings stated in Appendices B, C.2, and D, with checkpoint identifiers and source revisions provided with the released code. Every evaluation metric is defined explicitly, together with its thresholds, tolerances, and denominators: Appendix D.1 covers crystalline materials, Appendix D.2 covers molecular conformers, and Appendix C.1 covers protein–ligand complexes. The construction and filtering criteria of the EZBench benchmark are described in Appendix C.3. Together, the appendices specify the formulation, projection families, schedules, and evaluation protocols, and the released code and configuration files contain every numerical setting. Anonymized source code for AtomSteer, and the evaluation pipelines will be provided in an anonymized link, and we will release it publicly on GitHub upon publication.

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*, 630:493–500, 2024. doi: 10.1038/s41586-024-07487-w.

Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9:185, 2022. doi: 10.1038/s41597-022-01288-4.

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb–an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15(3):1652–1671, 03 2019. ISSN 1549-9618. doi: 10.1021/acs.jctc.8b01176. URL https://doi.org/10.1021/acs.jctc.8b01176.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 03 2024. ISSN 2041-6520. doi: 10.1039/d3sc04185a. URL https://doi.org/10.1039/d3sc04185a.

Siyuan Chen, Minghao Guo, Caoliwen Wang, Anka He Chen, Yikun Zhang, Jingjing Chai, Yin Yang, Wojciech Matusik, and Peter Yichen Chen. Physically valid biomolecular interaction modeling with gauss-seidel projection. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=sJABnBEYeh.

Mouyang Cheng, Wei-Liang Luo, Hao Tang, Bo-Wen Yu, Yongqiang Cheng, Wei-Wei Xie, Ju Li, Heather J. Kulik, and Mingda Li. Enhancing materials discovery with valence constrained design in generative modeling. *Nature computational science*, 2025. URL https://api.semanticscholar.org/CorpusID:280324280.

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained synthesis with projected diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 89307–89333, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/a28af221f2f70be183afc16797a56b91-Abstract-Conference.html.

Jacob K Christopher, Austin Seamann, Jingyi Cui, Sagar D Khare, and Ferdinando Fioretto. Constrained diffusion for protein design with hard structural constraints. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=kkvqVRu2Zy.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=qmCpJtFZra.

Octavian-Eugen Ganea, Lagnajit Pattanaik, Connor W. Coley, Regina Barzilay, Klavs F. Jensen, William H. Green, and Tommi S. Jaakkola. Geomol: torsional geometric generation of molecular 3d conformer ensembles. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, pp. 13, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

Thomas D. Goddard and Tristan I. Croll. Bad ligand chemistry in AlphaFold 3, OpenFold 3 and Boltz predictions. https://www.rbvi.ucsf.edu/chimerax/data/ligchem-feb2026/badchem.html, 2026. UCSF Resource for Biocomputing, Visualization, and Informatics.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. Et-flow: Equivariant flow-matching for molecular conformer generation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 128798–128824. Curran Associates, Inc., 2024. doi: 10.52202/079017-4091. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e8bd617e7dd0394ceadf37b4a7773179-Paper-Conference.pdf.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J. Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. *ArXiv*, abs/2311.16424, 2023. URL https://api.semanticscholar.org/CorpusID:265466093.

Philipp Höllmer, Thomas Egg, Maya M. Martirossyan, Eric Fuemmeler, Zeren Shui, Amit Gupta, Pawan Prakash, Adrian Roitberg, Mingjie Liu, George Karypis, Mark Transtrum, Richard G. Hennig, Ellad B. Tadmor, and Stefano Martiniani. Open materials generation with stochastic interpolants. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025.

Tatsuki Hori, Yoshitaka Moriwaki, and Ryuichiro Ishitani. Distance-restraint-guided diffusion models for sampling protein conformational changes and ligand dissociation pathways. *Journal of Chemical Theory and Computation*, 22(10):5259–5270, 05 2026. ISSN 1549-9618. doi: 10.1021/acs.jctc.6c00199. URL https://doi.org/10.1021/acs.jctc.6c00199.

Matthew K. Horton, Patrick Huck, Ruo Xi Yang, Jason M. Munro, Shyam Dwaraknath, Alex M. Ganose, Ryan S. Kingsbury, Mingjian Wen, Jimmy X. Shen, Tyler S. Mathis, Aaron D. Kaplan, Karlo Berket, Janosh Riebesell, Janine George, Andrew S. Rosen, Evan W. C. Spotte-Smith, Matthew J. McDermott, Orion A. Cohen, Alex Dunn, Matthew C. Kuner, Gian-Marco Rignanese, Guido Petretto, David Waroquiers, Sinead M. Griffin, Jeffrey B. Neaton, Daryl C. Chrzan, Mark Asta, Geoffroy Hautier, Shreyas Cholia, Gerbrand Ceder, Shyue Ping Ong, Anubhav Jain, and Kristin A. Persson. Accelerated data-driven materials science with the materials project. *Nature Materials*, 24:1522–1532, 2025. doi: 10.1038/s41563-025-02272-0.

Ryuichiro Ishitani and Yoshitaka Moriwaki. Improving stereochemical limitations in protein–ligand complex structure prediction. *ACS Omega*, 10(46):56075–56084, 11 2025. ISSN 2470-1343. doi: 10.1021/acsomega.5c07675. URL https://doi.org/10.1021/acsomega.5c07675.

Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 17464–17497. Curran Associates, Inc., 2023. doi: 10.52202/075280-0767. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/38b787fc530d0b31825827e2cc306656-Paper-Conference.pdf.

Rui Jiao, Wenbing Huang, Yu Liu, Deli Zhao, and Yang Liu. Space group constrained crystal generation. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 6836–6853, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/1ba1b2133ac3d3dda0ce5cd67dab8456-Paper-Conference.pdf.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022. doi: 10.52202/068431-1926. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/a98846e9d9cc01cfb87eb694d946ce6b-Paper-Conference.pdf.

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

Helen Lai, Tingyu Wang, Hassan Sirelkhatim, Joe Eaton, Howard Huang, Brad Rees, Ola Engkvist, Jon Paul Janet, Xiaoyun Wang, and Alessandro Tibo. Improving protein-ligand complex generation with force field guidance. *Journal of Cheminformatics*, 18(1), 2026.

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.

Xiaohan Lin, Yijie Xia, Yanheng Li, Yu-Peng Huang, Shuo Liu, Jun Zhang, and Yi Qin Gao. In-silico 3d molecular editing through physics-informed and preference-aligned generative foundation models. *Nature Communications*, 16:6043, 2025. doi: 10.1038/s41467-025-61323-x.

Lihang Liu, Shanzhuo Zhang, Yang Xue, Xianbin Ye, Kunrui Zhu, Yuxin Li, Yang Liu, Jie Gao, Wenlai Zhao, Hongkun Yu, Zhihua Wu, Xiaonan Zhang, and Xiaomin Fang. Technical report of helixfold3 for biomolecular structure prediction, 2024. URL https://arxiv.org/abs/2408.16975.

Xiaoshan Luo, Zhenyu Wang, Qingchang Wang, Xuechen Shao, Jian Lv, Lei Wang, Yanchao Wang, and Yanming Ma. CrystalFlow: A flow-based generative model for crystalline materials. *Nature Communications*, 16:9267, 2025. doi: 10.1038/s41467-025-64364-4.

Muhammad Luthfi, Adam J. Simpkin, Luc G. Elliott, Pornthep Sompornpisut, and Daniel J. Rigden. Physical implausibility of carbohydrate ligands in results of deep learning-based cofolding methods. *Journal of Chemical Information and Modeling*, 66(7):3456–3463, 04 2026. ISSN 1549-9596. doi: 10.1021/acs.jcim.5c03075. URL https://doi.org/10.1021/acs.jcim.5c03075.

Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M. Wood. Flowmm: generating materials with riemannian flow matching. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Filipp Nikitin, Dylan M. Anstine, Roman Zubatyuk, Saee Gopal Paliwal, and Olexandr Isayev. Scalable low-energy molecular conformer generation with quantum mechanical accuracy. *ChemRxiv*, 2026(0712), 2026. doi: 10.26434/chemrxiv-2025-k4h7v/v3. URL https://chemrxiv.org/doi/abs/10.26434/chemrxiv-2025-k4h7v/v3.

Ryotaro Okabe, Mouyang Cheng, Abhijatmedhi Chotrattanapituk, Manasi Mandal, Kiran Mak, Denisse Córdova Carrizales, Nguyen Tuan Hung, Xiang Fu, Bowen Han, Yao Wang, Weiwei Xie, Robert J. Cava, Tommi S. Jaakkola, Yongqiang Cheng, and Mingda Li. Structural constraint integration in a generative model for the discovery of quantum materials. *Nature Materials*, 25 (2):223–230, 2026.

Woon Bae Park, Jiyong Chung, Jaeyoung Jung, Keemin Sohn, Satendra Pal Singh, Myoungho Pyo, Namsoo Shin, and Kee-Sun Sohn. Classification of crystal structure using a convolutional neural network. *IUCrJ*, 4(4):486–494, 2017. doi: 10.1107/s205225251700714x. URL http://dx.doi.org/10.1107/S205225251700714X.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707. URL https://www.biorxiv.org/content/early/2025/06/18/2025.06.14.659707.

Alvaro Prat, Leo Zhang, Charlotte Deane, Yee Whye Teh, and Garrett M Morris. Sigmadock: Untwisting molecular docking with fragment-based SE(3) diffusion. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Vgm77U4ojX.

Zhuoran Qiao, Feizhi Ding, Thomas Dresselhaus, Mia A. Rosenfeld, Xiaotian Han, Owen Howell, Aniketh Iyengar, Stephen Opalenski, Anders S. Christensen, Sai Krishna Sirumalla, Frederick R. Manby, Thomas F. Miller III, and Matthew Welborn. Neuralplexer3: Accurate biomolecular complex structure prediction with flow models, 2024. URL https://arxiv.org/abs/2412.10743.

Yinuo Ren, Wenhao Gao, Lexing Ying, Grant M. Rotskoff, and Jiequn Han. Driftlite: Lightweight drift control for inference-time scaling of diffusion models, 2026. URL https://arxiv.org/abs/2509.21655.

Gianluca Scarpellini, Ron Shprints, Peter Holderrieth, Juno Nam, Pranav Murugan, Rafael Gómez-Bombarelli, Tommi Jaakkola, Maruan Al-Shedivat, Nicholas Matthew Boffi, and Avishek Joey Bose. Few-step cofolding with all-atom flow maps, 2026. URL https://arxiv.org/abs/2606.08375.

Kiyoung Seong, Sungsoo Ahn, Sehui Han, and Changyoung Park. Multimodal crystal flow: Any-to-any modality generation for unified crystal modeling. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=lKyD1ulXpY.

Jian Shen, Shengmin Zhou, and Xing Che. Fksfold: Improving alphafold3-type predictions of molecular glue–induced ternary complexes with feynman–kac–steered diffusion. *bioRxiv*, 2025. doi: 10.1101/2025.05.03.651455. URL https://www.biorxiv.org/content/early/2025/05/06/2025.05.03.651455.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen Mckeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55810–55827. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/singhal25b.html.

Yuta Suzuki, Hideitsu Hino, Takafumi Hawai, Kotaro Saito, Masato Kotsugi, and Kanta Ono. Symmetry prediction and knowledge discovery from x-ray diffraction patterns using an interpretable machine learning approach. *Scientific Reports*, 10:21790, 12 2020. doi: 10.1038/s41598-020-77474-4.

Chai Discovery Team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *bioRxiv*, 2024. doi: 10.1101/2024.10.10.615955. URL https://www.biorxiv.org/content/early/2024/10/11/2024.10.10.615955.

Genesis Research Team, Alejandro Dobles, Nina Jovic, Kenneth Leidal, Pranav Murugan, David C. Williams, Drausin Wulsin, Nate Gruver, Christina X. Ji, Korrawat Pruegsanusak, Gianluca Scarpellini, Ansh Sharma, Wojciech Swiderski, Andrea Bootsma, Richard Strong Bowen, Charlotte Chen, Jamin Chen, Marc André Dämgen, Benjamin DiFrancesco, J. D. Fishman, Alla Ivanova, Zach Kagin, David Li-Bland, Zuli Liu, Igor Morozov, Jeffrey Ouyang-Zhang, Frank C. Pickard IV, Kushal S. Shah, Ben Shor, Gabriel Monteiro da Silva, Roy Tal, Maxx Tessmer, Carl Tilbury, Cyr Vetcher, Daniel Zeng, Maruan Al-Shedivat, Aleksandra Faust, Evan N. Feinberg, Michael V. LeVine, and Matteus Pan. Pearl: A foundation model for placing every atom in the right location, 2025. URL https://arxiv.org/abs/2510.24670.

Protenix Team, Yuxuan Zhang, Chengyue Gong, Hanyu Zhang, Wenzhi Ma, Zhenyu Liu, Xinshi Chen, Jiaqi Guan, Lan Wang, and Wenzhi Xiao. Protenix-v1: Toward high-accuracy open-source biomolecular structure prediction. *bioRxiv*, 2026. doi: 10.64898/2026.02.05.703733. URL https://www.biorxiv.org/content/early/2026/02/09/2026.02.05.703733.

Atsushi Togo, Kohei Shinohara, and Isao Tanaka. Spglib: a software library for crystal symmetry search. *Science and Technology of Advanced Materials: Methods*, 4(1):2384822, 2024. doi: 10.1080/27660400.2024.2384822. URL https://doi.org/10.1080/27660400.2024.2384822.

Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Vincent Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. In *Advances in Neural Information Processing Systems*, 2025.

Vilya Research, Pascal Sturmfels, Milad Salem, Naozumi Hiranuma, Stephen Rettie, Xiaoliang Pan, Benjamin D. Sellers, Adam P. Moyer, Patrick J. Salveson, and Ivan Anishchanka. Vilya-1: An all-atom foundation model for macrocycle structure prediction and design. *arXiv preprint arXiv:2607.09998*, 2026.

David C. Williams and Neil Inala. Physics-informed generative model for drug-like molecule conformers. *Journal of Chemical Information and Modeling*, 64(8):2988–3007, 04 2024. ISSN 1549-9596. doi: 10.1021/acs.jcim.3c01816. URL https://doi.org/10.1021/acs.jcim.3c01816.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1 democratizing biomolecular interaction modeling. *bioRxiv*, 2025. doi: 10.1101/2024.11.19.624167. URL https://www.biorxiv.org/content/early/2025/05/06/2024.11.19.624167.

Brandon Wood, Misko Dzamba, Xiang Fu, Meng Gao, Muhammed Shuaibi, Luis Barroso-Luque, Kareem Abdelmaqsoud, Vahe Gharakhanyan, John Kitchin, Daniel Levine, Kyle Michel, Anuroop Sriram, Taco Cohen, Abhishek Das, Sushree Sahoo, Ammar Rizvi,

Zachary Ulissi, and Larry Zitnick. Uma: A family of universal models for atoms. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 129391–129427. Curran Associates, Inc., 2025. doi: 10.52202/085713-4310. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/bbf23e81b0ad7637fe9a731d0b676ca6-Paper-Conference.pdf.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models, 2024. URL https://arxiv.org/abs/2306.17775.

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_.

A MOLECULAR CONSTRAINTS AND THEIR PROJECTIONS

Molecular constraints specify bond geometry, nonbonded separation, planarity, and stereochemistry. We define them from the input molecular graph and apply them to the endpoint prediction. ET-Flow and Boltz-2 use the same chemical requirements and projections (Appendices B.2 and C.2).

A.1 BOND GEOMETRY, STERIC SEPARATION, AND PLANARITY

Bond lengths. A covalent bond admits a limited range of lengths determined by its atoms and bond type. For a bonded pair (i, j), let $d_{ij} = \|x_i - x_j\|_2$ and let $[\bar{\ell}_{ij}, \bar{u}_{ij}]$ be its allowed distance interval after the configured tolerance is applied. The lower and upper violations are

$$r_{ij}^{\text{distance}}(x) = \begin{pmatrix} [\bar{\ell}_{ij} - d_{ij}]_+ \\ [d_{ij} - \bar{u}_{ij}]_+ \end{pmatrix}, \qquad [z]_+ = \max(z, 0). \tag{5}$$

Both entries vanish within the interval. These intervals are constructed from connectivity and atom types, allowing bond lengths to vary within chemically admissible ranges.

Bond angles. For two bonds (i, j) and (j, k), the distance between atoms i and k constrains the angle at j through

$$d_{ik}^2 = d_{ij}^2 + d_{jk}^2 - 2d_{ij}d_{jk}\cos\theta_{ijk}. \tag{6}$$

We apply Eq. (5) to these 1–3 pairs using their graph-derived distance bounds. Together with the bond-length residuals, these constraints retain the local molecular shape while allowing variation within the specified intervals.

Intramolecular steric clashes. Atoms that are not close covalent neighbors must remain sufficiently separated. For each eligible nonbonded pair, the graph-derived exclusion distance $\bar{\ell}_{ij}^{\text{nb}}$ gives the residual

$$r_{ij}^{\text{clash}}(x) = [\bar{\ell}_{ij}^{\text{nb}} - d_{ij}]_+. \tag{7}$$

This residual acts only when atoms approach more closely than their permitted separation. Bonded and 1–3 pairs are handled by their own distance constraints. Protein–ligand and intercomponent exclusions are defined separately in Appendix C.2.1, where atom-type radii and the surrounding protein determine the relevant pairs.

Ring and double bond planarity. Aromatic rings and the local trigonal geometry around a double bond require approximately coplanar atoms. For a specified group S, define its centroid $\bar{x}_S$ and the normal n_S of its least-squares plane:

$$\begin{aligned} \bar{x}_S &= \frac{1}{|S|}\sum_{i\in S} x_i, \qquad & n_S &\in \arg\min_{\|n\|_2=1} \sum_{i\in S} [(x_i - \bar{x}_S)^\top n]^2, \\ r_i^{\text{plane}} &= [|(x_i - \bar{x}_S)^\top n_S| - \varepsilon_S]_+. \end{aligned} \tag{8}$$

The residual measures displacement beyond the permitted plane thickness ε_S. Ring and double bond groups are specified separately, so a planarity requirement is applied only to the chemically designated atoms.

A.2 DOUBLE BOND STEREOCHEMISTRY AND TETRAHEDRAL CHIRALITY

***E/Z* configuration.** For a stereogenic double bond $j = k$, the molecular graph specifies which substituents define its orientation. The graph-ordered tuple (i, j, k, l) has signed dihedral ϕ_{ijkl}. Writing $u = (x_k - x_j)/\|x_k - x_j\|_2$, $v = (x_i - x_j) - [(x_i - x_j)^\top u]u$, and $w = (x_l - x_k) - [(x_l - x_k)^\top u]u$, we use

$$\phi_{ijkl} = \text{atan2}[(u \times v)^\top w, v^\top w], \qquad e_{ijkl} = \text{atan2}[\sin(\phi_{ijkl} - \phi_\star), \cos(\phi_{ijkl} - \phi_\star)]. \tag{9}$$

In the graph-oriented convention, $\phi_\star = \pi$ denotes *E* and $\phi_\star = 0$ denotes *Z*. The signed error e_{ijkl} is periodic, so the equivalent trans angles $-\pi$ and π have the same target. Our central-dihedral refinement uses $r^{\text{EZ}} = e_{ijkl}/s_\phi$ with angular scale s_ϕ. An admissible band of half-width b_ϕ instead has violation $[|e_{ijkl}| - b_\phi]_+$. The center residual controls torsional accuracy, while the band identifies whether the requested stereochemical configuration is satisfied.

Tetrahedral handedness. A stereocenter's ordered substituents determine its *R*/*S* configuration. For four graph-ordered positions, the oriented volume is proportional to

$$V_{ijkl} = (x_i - x_l)^\top [(x_j - x_l) \times (x_k - x_l)]. \tag{10}$$

Its sign changes under reflection. The graph ordering determines which sign agrees with the input stereochemistry. For local chirality correction we use the corresponding signed-improper margin $m_i = s_i \phi_i$, where ϕ_i is the signed improper dihedral at center i and $s_i \in \{-1, 1\}$ is the graph-derived orientation sign. The residual

$$r_i^{\mathrm{RS}}(x) = [\mu - m_i(x)]_+ \tag{11}$$

requires the prescribed orientation with a positive signed-angle margin μ. In ET-Flow, its graph-ordered four-substituent dihedral margins select the conformer's global parity after sampling (Eq. (15)).

A.3 Structure-dependent molecular projections

The molecular implementations use two recurring projection families. ChemGN reduces local chemical residuals through Cartesian or bond-tangent updates. Rigid-fragment and rigid-body projections move connected fragments or complete components while preserving their internal geometry. Global parity selection, a final conformer-specific step that chooses between a conformer and its mirror image, is a separate operation.

Cartesian ChemGN. ChemGN is the umbrella local chemical projection family. Its Cartesian form reduces the active residuals by changing selected Cartesian coordinates. Let q contain those coordinates and $J = \partial r / \partial q$ be the residual Jacobian. A damped linearized correction is

$$\Delta q = \arg\min_d \frac{1}{2}\|r + Jd\|_2^2 + \frac{\epsilon}{2}\|d\|_2^2 = -J^\top (JJ^\top + \epsilon I)^{-1} r. \tag{12}$$

Residual normalization is specified in Appendices B.2 and C.2.1. We apply the update separately to each residual family over a bounded set of active residuals, and a blend factor β and movement setting $\Delta_{\max}$ bound the resulting correction. This is the local residual-reduction step underlying the molecular instantiations of Eq. (3).

Bond-tangent ChemGN for local *R*/*S*. Within a protein complex, we correct ligand chirality while protecting its bond geometry. This is the ChemGN update of Eq. (12) with a tangent-restricted Jacobian. Let B be the Jacobian of the protected bond distances. We restrict the chirality Jacobian using

$$T_{\eta_{\tan}} = I - B^\top (BB^\top + \eta_{\tan} I)^{-1} B, \qquad \bar{J} = J_{\mathrm{RS}} T_{\eta_{\tan}}. \tag{13}$$

The Gauss–Newton solve uses $\bar{J}$ to reduce Eq. (11). This regularized tangent restriction limits first-order bond changes. An update is accepted only if it keeps already-correct *E*/*Z* assignments and *R*/*S* signs and stays within the displacement bounds.

Rigid-fragment *E*/*Z* projection. Correcting a double bond orientation can require moving an entire substituent rather than only the four atoms used to measure its dihedral. Let $\mathcal{W}$ be an eligible connected fragment, with bond axis u through p. We apply

$$x_i' = \begin{cases} p + R_u(\vartheta)(x_i - p), & i \in \mathcal{W}, \\ x_i, & i \notin \mathcal{W}. \end{cases} \tag{14}$$

The rotation preserves every distance within $\mathcal{W}$. Graph connectivity, ring closures, and covalent attachments determine which fragments can rotate. The rigid-fragment projection acts in two steps: a flip selects the rotation into the requested stereochemical basin, and a refinement adjusts the torsion with bounded rotations of the same fragment. Cartesian ChemGN then reduces the remaining local geometric residuals.

Rigid-body component projection. A ligand or protein component can reduce overlap with its environment through $x_i' = Rx_i + b$, with $(R, b) \in SE(3)$. Proper rigid motion preserves the component's internal distances and stereochemistry. We optimize its translation and rotation against the intercomponent residuals in Appendix C.2.1. Covalently connected chains move together, preserving their internal connections.

Table 7: **AtomSteer instantiations on the three backbones.** Every task runs Algorithm 1. Percentages give the share of sampling steps completed when a projection acts (FlowMM 1,000, ET-Flow 50, and Boltz-2 200 steps); 100% and "terminal" denote the final state, and "to end" includes the final sampling step. Molecular projections are listed in their per-step order. A projection runs only when its constraint applies to the structure, and coordinates without an accepted correction keep the native update. FlowMM keeps its native velocity scaling, $Q(A, F, \ell) = (1, 1 + 5t, 1 + 5t)$ for de novo and motif-constrained generation and $(1, 1 + 10t, 1)$ for both CSP variants.

Task	Endpoint prediction	Active projection families	When projections act (percentage of steps)	Continuation
FlowMM de novo	Velocity extrapolation; native terminal state at 100%	Crystal symmetry: inferred orbits and lattice metric	Every 2.5% over the last 20% (80% to 100%)	Endpoint displacement (25)
FlowMM CSP	Velocity extrapolation; native terminal state at 100%	Crystal symmetry: lattice-metric template (inferred or supplied) and inferred orbits	Composition-only: lattice and coordinates at 90% and 100%. Lattice-informed: lattice every 1%, coordinates at 90% and 100%	Endpoint displacement (25)
FlowMM motif	Velocity extrapolation	Crystal motif: species, positions, and in-plane cell	Every step	Initial-state interpolation (26)
ET-Flow conformers	Velocity extrapolation	ChemGN; rigid-fragment *E*/*Z*; global parity selection	rigid-fragment flip 16% to 80%; refinement 32% to end; Cartesian ChemGN 16% to end; terminal ChemGN, then parity selection	Initial-state interpolation (28)
Boltz-2 complexes	Denoised EDM estimate	ChemGN (Cartesian and bond-tangent *R*/*S*); rigid-fragment *E*/*Z*; rigid-body ligand and protein components	rigid-fragment flip 15% to 80%; refinement 32.5% to end; Cartesian ChemGN 15% to end; ligand $SE(3)$ 67.5% to end; protein-component $SE(3)$ 60% to end; bond-tangent ChemGN (*R*/*S*) every step; terminal ligand, protein-component, and *R*/*S* projections	EDM update (29)

Global conformer parity. For an isolated conformer, we compare its existing orientation with a reflection through a plane passing through its centroid:

$$x_i' = \bar{x} + H(x_i - \bar{x}), \qquad H = \mathrm{diag}(1, 1, -1). \tag{15}$$

This changes global handedness while preserving pairwise distances, bond angles, and planarity magnitudes. We keep the orientation with more correctly signed stereocenters under the graph-oriented dihedral margins and use the margin magnitudes to break ties. This projection chooses global parity rather than independently changing individual stereocenters.

B Pre-trained backbone integration and sampling schedules

All three backbones run one AtomSteer loop (Algorithm 1). At each step, one model evaluation gives the endpoint prediction, the projections scheduled for that step apply bounded corrections that must pass their acceptance checks, and the native update continues from the projected endpoint. Table 7 lists, for each task, the endpoint prediction, the active projection families, when they act as a share of the sampling trajectory, and the continuation rule. AtomSteer keeps the sampler of each pre-trained backbone as originally released: ET-Flow uses 50 flow-matching steps, FlowMM 1,000 steps, and Boltz-2 200 diffusion steps. Because the projections are adapted to these native samplers, their schedules differ between backbones. The subsections below specify the FlowMM and ET-Flow implementations, and Appendix C.2 specifies the Boltz-2 implementation. Exact numerical settings are provided with the released code.

Algorithm 1 AtomSteer sampling with a pre-trained backbone (one trajectory)

1: Initialize the native sampler state, retaining the initial state when the continuation uses it.
2: **for** each native sampling step from t to t' **do**
3: Evaluate the model once to form the native update $x_{t'}^{\mathrm{nat}}$ and the endpoint prediction $\hat{x}_t^{\mathrm{end}}$ (Eq. (4)).
4: Set $\tilde{x}_t^{\mathrm{end}} \leftarrow \hat{x}_t^{\mathrm{end}}$.
5: **for** each projection scheduled at this step, in order (Table 7) **do**
6: Propose a bounded correction of $\tilde{x}_t^{\mathrm{end}}$ and keep it only if it passes the projection's acceptance checks.
7: **end for**
8: **if** no correction was kept **then**
9: $x_{t'} \leftarrow x_{t'}^{\mathrm{nat}}$.
10: **else**
11: $x_{t'} \leftarrow$ native update from x_t with $\tilde{x}_t^{\mathrm{end}}$ in place of $\hat{x}_t^{\mathrm{end}}$ (Section 3.2).
12: **end if**
13: **end for**
14: Apply the scheduled terminal projections to the final state with the same bounds and acceptance checks.

B.1 FlowMM: crystal generation and structure prediction

Endpoint prediction. We use FlowMM as a pre-trained backbone for symmetry and motif constraints in crystal generation (Miller et al., 2024). De novo and motif-constrained generation use the configuration-matched unconditional FlowMM checkpoint, and CSP uses the configuration-matched composition-conditioned checkpoint; checkpoint identifiers are provided with the released code. The model learns a Riemannian vector field over atom types, fractional positions, and lattice parameters, with sampling state $x_t = (A_t, F_t, \ell_t)$ and $F_t \in \mathbb{T}^{N\times 3}$. Atom types use continuous analog bits in de novo and motif-constrained generation, and composition is fixed in CSP. The native lattice chart ℓ holds the three cell lengths and transformed cell angles. Decoding gives the physical row-vector cell L and metric $G = LL^\top$; crystallographic projections use L and G, and sampling updates use ℓ.

All tasks use FlowMM's native sampler with 1,000 Euler steps. At time t and next time τ, one model evaluation gives the raw field, the native update, and the endpoint prediction,

$$\begin{aligned} v_{\mathrm{raw}} &= v_\theta(x_t, t), & x_\tau^{\mathrm{nat}} &= \mathrm{projx}[x_t + (\tau - t)Q(t)v_{\mathrm{raw}}], \\ \hat{x}_t^{\mathrm{end}} &= \mathrm{projx}[x_t + (1-t)v_{\mathrm{raw}}] & & \text{endpoint extrapolation.} \end{aligned} \tag{16}$$

where $Q(t)$ is FlowMM's native velocity scaling (Table 7). At the final step, de novo generation and CSP take the native terminal state x_1^{nat} as the endpoint prediction, and motif-constrained generation keeps the raw-field extrapolation. For fractional coordinates, $\Delta_{\mathbb{T}}(F_1, F_2) = d - \mathrm{round}(d)$ with $d = F_2 - F_1$ is the componentwise minimum-image displacement, and $\mathrm{wrap}(u) = u - \lfloor u \rfloor$.

B.1.1 Crystal symmetry projections

Crystal symmetry projection couples two requirements: symmetry-related fractional coordinates and a compatible lattice metric.

Fractional-coordinate symmetry. A crystal symmetry maps atoms onto sites occupied by the same species, including sites in neighboring unit cells. A realization specifies fractional operations (R_g, b_g), species-preserving permutations π_g, and a cell setting. Writing each fractional position f_i as a column, its residual is

$$r_{F,g,i} = \Delta_{\mathbb{T}}(f_{\pi_g(i)}, R_g f_i + b_g). \tag{17}$$

The residual vanishes when each transformed atom coincides with its same-species partner modulo a lattice translation.

Table 8: **Lattice-metric constraints.** Each listed quantity vanishes for a cell in the stated setting.

Metric family	Entries of the constraint residual
Triclinic	No additional metric equalities
Monoclinic (b unique)	$G_{12},\ G_{23}$
Orthorhombic	$G_{12},\ G_{13},\ G_{23}$
Tetragonal (c unique)	$G_{12},\ G_{13},\ G_{23},\ G_{11}-G_{22}$
Cubic	$G_{12},\ G_{13},\ G_{23},\ G_{11}-G_{22},\ G_{22}-G_{33}$
Hexagonal (c unique)	$G_{13},\ G_{23},\ G_{11}-G_{22},\ 2G_{12}+G_{11}$
Rhombohedral	$G_{11}-G_{22},\ G_{22}-G_{33},\ G_{12}-G_{13},\ G_{13}-G_{23}$

Lattice-metric symmetry. The cell must make each fractional rotation a distance-preserving physical operation. Its metric records the lengths and angles of the lattice vectors:

$$G = LL^\top = \begin{pmatrix} a^2 & ab\cos\gamma & ac\cos\beta \\ ab\cos\gamma & b^2 & bc\cos\alpha \\ ac\cos\beta & bc\cos\alpha & c^2 \end{pmatrix}, \qquad r_{G,g} = R_g^\top G R_g - G. \tag{18}$$

Zero metric residual means that the symmetry preserves distances in the periodic structure. Atomic and metric residuals are imposed together, so the selected symmetry is compatible with both the positions and the cell.

Hierarchy-aware symmetry selection infers compatible realizations from the endpoint prediction. Symmetry detection over a range of tolerances and cell settings enumerates the realizations that the endpoint prediction approximately satisfies. Selection starts from the feasible realization that needs the least projection work, follows minimal-supergroup relations to the most symmetric feasible parent, and uses projection work only to choose between parents that are not related by inclusion.

Within a chosen atomic orbit, the predicted images determine a representative q_0. Operations that leave the representative on the same periodic site give affine constraints $Bq = b$ after fixing their integer translations. The representative and its symmetry-related sites are projected by

$$q^\star = q_0 + B^\dagger(b - Bq_0), \qquad f^\star_{\pi_g(i)} = R_g q^\star + b_g \pmod{\mathbb{Z}^3}. \tag{19}$$

For de novo generation, the lattice projector averages the predicted metric $\hat{G} = \hat{L}\hat{L}^\top$ over the selected rotation set $\mathcal{S}_a$:

$$G^{\mathrm{av}} = \frac{1}{|\mathcal{S}_a|}\sum_{R\in\mathcal{S}_a} R^\top \hat{G} R, \qquad G^\star = \left(\frac{\det \hat{G}}{\det G^{\mathrm{av}}}\right)^{1/3} G^{\mathrm{av}}. \tag{20}$$

The average satisfies the selected metric symmetry, and the determinant rescaling retains the predicted endpoint volume. A Procrustes lift constructs a cell with metric $G^\star$ near the predicted orientation. Atom types and order remain fixed during these symmetry projections. If no eligible realization is found, the native update is used.

Lattice-metric constraints in CSP. CSP projects the metric in a conventional setting S_{setting} onto a lattice-metric template $\mathcal{T}$:

$$\begin{aligned} &G_c = (S_{\mathrm{setting}}\hat{L})(S_{\mathrm{setting}}\hat{L})^\top, && G_c^\star = P_{\mathcal{T}}(G_c), \\ &L^\star = \mathrm{VolRestore}_{\hat{L}}[S_{\mathrm{setting}}^{-1}\,\mathrm{Lift}(G_c^\star)]. \end{aligned} \tag{21}$$

The lift uses a positive-definite factorization in the selected setting. Table 8 gives the zero-residual metric relations in representative axis conventions; the selected template also fixes the axis convention used for projection.

The template projector $P_{\mathcal{T}}$ is the least-squares projection of the independent metric entries onto these relations: it sets the required entries to zero and fits the common parameter of each group of related entries by least squares. Positive-definite factorization and volume restoration then give the projected cell.

Composition-only and lattice-informed CSP. In composition-only CSP, hierarchy-aware symmetry selection infers the metric template, its setting, and the orbit realization from the endpoint prediction, then corrects the lattice and the fractional coordinates.

Lattice-informed CSP receives a reference-derived metric template, which fixes one of the seven lattice systems in Table 8 and its axis convention, and an integer setting transform to the conventional cell. The lattice-informed setting also uses the reference point-group rotations to check metric compatibility. FlowMM generates the numerical lengths, angles, volume, and fractional coordinates. The lattice is projected onto this template, and hierarchy-aware symmetry selection infers coordinate symmetry from the lattice-corrected endpoint prediction and projects only the fractional coordinates; an incompatible coordinate projection is skipped.

B.1.2 Motif projections

Motif projections complete the crystal symmetry and motif projection family of Table 1. We apply them to SCIGEN's requested Archimedean lattice (AL) task (Okabe et al., 2026). Each condition specifies a motif, atomic species, and nearest-neighbor distance d. The constraint fixes the designated atoms' species, their periodic arrangement, and the cell's in-plane dimensions. It leaves the remaining atoms and the other cell parameters available to the generator.

Specified atomic species. For requested atomic number Z, let a_Z be its FlowMM analog-bit encoding. The bth bit has value $a_{Z,b} = s_{\mathrm{bit}}[2\,\mathrm{bit}_b(Z-1)-1]$, where s_{bit} is the model's bit scale. For each designated site $i \in S$, the species residual is $r_{A,i} = A_i - a_Z$. Projection assigns $A_i = a_Z$ on these sites.

Periodic atomic pattern. The basal fractional patterns are $\{(0,0)\}$ for triangular, $\{(1/3,2/3),(2/3,1/3)\}$ for honeycomb, and $\{(0,0),(1/2,0),(0,1/2)\}$ for Kagome. Let m_i be these sites embedded in a common fractional plane. The positional residual and the selected common translation are

$$r^{\mathrm{AL}}_{F,i}(F,\zeta) = \Delta_{\mathbb{T}}(f_i, m_i + \zeta), \qquad \zeta^\star \in \arg\min_{\zeta\in\mathbb{T}^3} \sum_{i\in S} \|r^{\mathrm{AL}}_{F,i}(F,\zeta)\|_2^2. \tag{22}$$

We compute the componentwise periodic least-squares translation and set $f_i^\star = \mathrm{wrap}(m_i + \zeta^\star)$. This preserves the specified pattern while placing it near the predicted atoms.

Absolute lattice scale. The cell must realize the requested nearest-neighbor separation. For $s = 1, \sqrt{3}, 2$ in the triangular, honeycomb, and Kagome cases, respectively, the cell residual is

$$r^{\mathrm{AL}}_L(L) = (a - sd,\ b - sd,\ \gamma - 2\pi/3). \tag{23}$$

The projection sets $a = b = sd$ and $\gamma = 120^\circ$, retaining c, α, β. Together, the species, position, and cell assignments satisfy the AL requirements at the projected endpoint.

B.1.3 Projection schedule

Table 7 lists when each projection acts.

B.1.4 Bounded corrections and acceptance

Periodic contacts depend on both the fractional coordinates and the physical cell, so each trial lattice is decoded before it is checked. De novo generation and CSP first rescale the trial cell isotropically to the volume of the same-step native cell and re-encode it in the native chart. For contact checks, the shortest Cartesian separation in the decoded cell is

$$d^{\mathrm{PBC}}_{ij} = \min_{n\in\mathbb{Z}^3} \|(f_j - f_i - n)L\|_2, \qquad r^{\mathrm{contact}}_{ij} = [d_0 - d^{\mathrm{PBC}}_{ij}]_+, \tag{24}$$

with minimum separation d_0; for self-images, $i = j$ and $n = 0$ is excluded. Every trial cell must be finite, exceed a minimum volume, and remain well conditioned, and in de novo generation and CSP its conditioning must stay close to that of the native cell.

Each correction is bounded. De novo generation bounds the endpoint coordinate displacement, and CSP additionally bounds the coordinate change at the next sampling state and the stretch of the trial metric relative to the native metric. Every correction must pass the periodic-contact check at separation d_0, CSP also limits any decrease of the minimum interatomic distance, and every trial cell must re-encode faithfully in FlowMM's lattice chart. A failed check rejects the correction, and the affected coordinates keep their native update. De novo generation accepts or rejects the coordinate and lattice corrections together, whereas in CSP a rejected coordinate correction does not undo an accepted lattice correction. Motif-constrained generation keeps the requested absolute cell scale instead of restoring the native volume. Exact settings for FlowMM are provided with the released code.

B.1.5 NATIVE CONTINUATION

After endpoint projection, the trial update changes only the selected coordinates. In de novo generation and CSP, the endpoint displacement enters the native update,

$$\begin{aligned} A_\tau^{\mathrm{trial}} &= A_\tau^{\mathrm{nat}} + \tau M_A(\tilde{A}_t^{\mathrm{end}} - \hat{A}_t^{\mathrm{end}}), \\ F_\tau^{\mathrm{trial}} &= \mathrm{wrap}\Big[F_\tau^{\mathrm{nat}} + \tau M_F \Delta_{\mathbb{T}}(\hat{F}_t^{\mathrm{end}}, \tilde{F}_t^{\mathrm{end}})\Big], \\ \ell_\tau^{\mathrm{trial}} &= \ell_\tau^{\mathrm{nat}} + \tau M_\ell(\tilde{\ell}_t^{\mathrm{end}} - \hat{\ell}_t^{\mathrm{end}}). \end{aligned} \tag{25}$$

where the masks M_A, M_F, and M_ℓ select the atoms and lattice entries corrected by each constraint and all other entries are copied from the native update.

Motif-constrained generation instead retains the initial state $x_0 = (A_0, F_0, \ell_0)$ and interpolates toward the projected endpoint. With $g(F, H; s) = \mathrm{wrap}[F + s\,\Delta_{\mathbb{T}}(F, H)]$, a changed endpoint gives the trial state

$$\begin{aligned} A_\tau^{\mathrm{trial}} &= A_\tau^{\mathrm{nat}} + \tau\big[(1-\tau)A_0 + \tau\tilde{A}_t^{\mathrm{end}} - A_\tau^{\mathrm{nat}}\big], \\ F_\tau^{\mathrm{trial}} &= g\Big(F_\tau^{\mathrm{nat}}, g(F_0, \tilde{F}_t^{\mathrm{end}}; \tau); \tau\Big), \\ \ell_\tau^{\mathrm{trial}} &= \ell_\tau^{\mathrm{nat}} + \tau\big[(1-\tau)\ell_0 + \tau\tilde{\ell}_t^{\mathrm{end}} - \ell_\tau^{\mathrm{nat}}\big]. \end{aligned} \tag{26}$$

The final scheduled projection uses this continuation at $\tau = 1$. An unchanged endpoint or a rejected trial keeps the native update.

B.2 ET-FLOW: MOLECULAR CONFORMER GENERATION

Endpoint prediction. We use ET-Flow as a pre-trained backbone for molecular constraints (Hassan et al., 2024). Every ET-Flow row uses the released GEOM-Drugs checkpoint `drugs-o3`, which Hassan et al. (2024) evaluate as ET-Flow and ET-Flow-SS, with the same molecular inputs and reported conformer budgets. ODE50 and SS50 use the released deterministic and stochastic samplers and the released post hoc chirality correction, which AtomSteer replaces with global parity selection. The model generates Cartesian conformers conditioned on molecular topology, transporting an initial state $x_0 \sim \rho_0$ toward the conformer distribution at $t = 1$. AtomSteer uses 50 Euler steps. At time t and next time τ, the velocity gives the endpoint prediction and the native update:

$$\hat{x}_t^{\mathrm{end}} = x_t + (1-t)v_\theta(x_t, t), \qquad x_\tau^{\mathrm{nat}} = x_t + (\tau - t)v_\theta(x_t, t). \tag{27}$$

Molecular geometry constraints. The input graph defines the conformer's covalent connectivity and stereochemistry. From it we construct bond-distance intervals, 1–3 distance intervals that constrain bond angles through Eq. (6), nonbonded exclusions, and aromatic and double bond planar groups, with the residuals of Eqs. (5)–(8) under graph-derived bounds and residual scaling. Cartesian ChemGN reduces these residuals (Appendix A.3).

Stereochemical constraints. The molecular graph supplies the ordered substituents and target configuration of each stereogenic double bond. The central-dihedral residual is $r^{\mathrm{EZ}} = e/s_\phi$, with the wrapped angular error e of Eq. (9) and angular scale s_ϕ. The rigid-fragment *E/Z* projection rotates a graph-connected substituent (Eq. (14)): it flips a double bond that lies in the wrong *E/Z* class into the target basin and refines the same fragment with bounded rotations, and Cartesian ChemGN then reduces the remaining *E/Z* and local-geometry residuals. Both steps use the same

rotation rules and bounds as for Boltz-2. Global parity selection sets tetrahedral handedness from the graph-oriented signed dihedral margins of Appendix A.2.

Projection schedule. Table 7 lists when each projection acts. At each scheduled step, the rigid-fragment *E*/*Z* projections act first, followed by one Cartesian ChemGN update (Eq. (12)) for each of the *E*/*Z* center, bond length, 1–3 distance, intramolecular clash, and planarity families, in that order. After the last flow update, a terminal stage repeats the applicable Cartesian ChemGN families, and global parity selection then compares the conformer with its centroid-plane reflection (Eq. (15)). AtomSteer generates one trajectory per conformer.

Bounded corrections and acceptance. A rigid-fragment flip is accepted only if the double bond lands in a narrow band around its target, and a refinement move is accepted only if it worsens none of the bond, 1–3 distance, clash, planarity, *E*/*Z* center, and *R*/*S* residuals. A blend factor and a movement setting bound each Cartesian ChemGN update, and the terminal stage bounds the total displacement and checks each residual family. Exact settings for ET-Flow are provided with the released code.

Native continuation. AtomSteer retains the initial state x_0. When projection changes the endpoint, the next state interpolates from x_0 toward the projected endpoint; otherwise the native Euler update is used:

$$x_\tau = \begin{cases} (1-\tau)x_0 + \tau\tilde{x}_t^{\mathrm{end}}, & \tilde{x}_t^{\mathrm{end}} \neq \hat{x}_t^{\mathrm{end}}, \\ x_\tau^{\mathrm{nat}}, & \tilde{x}_t^{\mathrm{end}} = \hat{x}_t^{\mathrm{end}}. \end{cases} \tag{28}$$

C Protein–ligand generation

C.1 Evaluation metrics

PoseBusters validity. We run PoseBusters 0.6.5 in `redock` mode (Buttenschoen et al., 2024), conditioning the intermolecular checks on each method's own predicted complex rather than on the deposited receptor. PB-valid is the conjunction of the 27 checks of that profile, with two qualifications. PoseBusters' internal `rmsd_le_2A` check is excluded because it compares predicted and crystal ligands with no alignment; docking accuracy is evaluated separately after pocket alignment. And because a co-folding prediction contains no waters, the two water checks pass vacuously for every method, as do the four cofactor checks on any target whose prediction contains no cofactor. PB-valid is therefore a comparative validity endpoint under a fixed protocol.

Docking and joint success. Docking RMSD is the symmetry-corrected ligand heavy-atom RMSD to the crystal ligand after Cα alignment of the predicted and experimental 10 Å binding pockets. Docking success requires RMSD < 2 Å, strictly. Top-1 is the pose with the highest native Boltz-2 confidence among the five diffusion samples, where the native score is $(4\,\mathrm{pLDDT}_{\mathrm{complex}} + \mathrm{ipTM})/5$. For AtomSteer this confidence is recomputed on the final projected coordinates before selection, so the reported pose is the one the deployed pipeline would return. Success@1 requires that single confidence-selected pose to be both PB-valid and to have an RMSD < 2 Å. Success@5 requires at least one of the five returned poses to satisfy both conditions simultaneously; it is an oracle any-of-five statistic, not a deployable selector. All other single-pose rows of Table 6 are computed on the same confidence-selected pose.

Stereochemistry. For each evaluable stereogenic double bond we take the wrapped absolute difference between the predicted and crystal central dihedral, average over the bonds of one ligand, and report the distribution's mean, median and 90th percentile in degrees. A double bond is evaluable when RDKit assigns it an *E*/*Z* configuration after all hydrogens are removed and each of its two atoms carries a heavy-atom substituent, the same rule we apply to molecular conformers (Appendix D.2.2). Both stereochemistry success rows instead use a stricter construction in which a ligand passes only if *every* stereogenic double bond lies within 10° of the experimental reference, equivalently if its per-bond maximum does (*E*/*Z* row), or only if *every* evaluable tetrahedral center matches the reference assignment (*R*/*S* row). A tetrahedral center is evaluable

when the input declares its *R*/*S* label and an *R*/*S* label can be assigned in the predicted pose. Success@1+stereo is therefore exactly the conjunction of Success@1 with those two rows. The distinction from the per-ligand-mean reading is meaningful, since 290 of the 1,103 EZBench targets carry more than one stereogenic double bond and 371 of the 538 *R*/*S* targets carry more than one chiral center, and it lowers every method's level while widening AtomSteer's margin: 59.6, 59.5 and 79.1% with +19.7 points on *E*/*Z* against 62.9, 65.1 and 82.3% with +17.2, and 51.7, 85.3 and 85.1% on *R*/*S* against 75.4, 97.4 and 97.4%. We adopt the strict form since PoseBusters' own `double_bond_stereochemistry` and `tetrahedral_chirality` checks are likewise all-or-nothing per ligand. Discrete descriptor accuracy is secondary and saturates above 99% for most steered methods, because that PoseBusters check compares the InChI stereochemical layer and so resolves which side of the double bond plane each substituent occupies but not the local geometry and the angle at which it sits.

Independent structural scoring. As a preservation check we rescore the same coordinates with OpenStructure 2.12.0, which supplies lDDT-PLI, complex lDDT, backbone lDDT, TM-score, GDT-TS and Cα RMSD for the protein. This suite is independent of our own evaluator and of PoseBusters, and it contains no stereochemical term, so it measures whether the constraint layer degrades the folded structure or the interface, not whether it improves stereochemistry. Across the three methods, complex lDDT, backbone lDDT, TM-score, and Cα RMSD agree to within 0.004, so AtomSteer preserves the quality of the folded structure.

Denominators. Each table uses the maximal complete-case intersection over exactly the methods it displays, and each metric family carries its own intersection, because a ligand may have *E*/*Z* bonds, *R*/*S* centers, both, or neither. Targets without the relevant stereochemical feature are excluded from the stereochemistry counts rather than counted as successes. A target enters a count only if the metric is computable for every displayed method.

C.2 Boltz-2: protein–ligand generation and comparison methods

Protein–ligand generation with Boltz-2. We use Boltz-2 as a pre-trained backbone for applying molecular and intermolecular constraints to protein–ligand complexes (Passaro et al., 2025). The denoiser predicts all-atom coordinates from a noisy complex. Atom-to-token mappings, bonds, stereochemical labels, and component memberships specify the relations used by AtomSteer. AtomSteer reads the *E*/*Z* and *R*/*S* labels from the input isomeric SMILES, which is identical for all methods (crystal-verified for EZBench and taken from the official starting-ligand SDF for the PoseBusters benchmark). We retain Boltz-2 v2.2.1 with 200 denoising steps, three recycling iterations, FP32 arithmetic, seed 42, and five returned diffusion samples per target, with the top-1 sample selected by native Boltz confidence. Each trajectory uses one particle and AtomSteer projections in place of native potential guidance. No method has access to the deposited coordinates, to any evaluation metric, or to the reference stereochemical labels at any point during generation or selection. Boltz-2 follows EDM (Karras et al., 2022), with unscaled noisy coordinates $x_\sigma = x_0 + \sigma\epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, and σ decreasing to 0 during sampling. Its denoised estimate $\hat{x}^{\mathrm{EDM}}_{0,k} = D_\theta(x_{\hat{\sigma}_k}, \hat{\sigma}_k)$ is the endpoint prediction $\hat{x}^{\mathrm{end}}_{\hat{\sigma}_k}$.

Native continuation. Boltz-2 keeps its native schedule of $K = 200$ noise levels decreasing to $\sigma_K = 0$, its stochastic churn, which raises the noise level from σ_k to $\hat{\sigma}_k \geq \sigma_k$ before each denoiser call, and its coordinate augmentation and centering. AtomSteer projects $\hat{x}^{\mathrm{EDM}}_{0,k}$ to $\tilde{x}^{\mathrm{EDM}}_{0,k}$. Native reverse alignment then maps the noisy coordinates to this prediction's frame, $\bar{x}_{\hat{\sigma}_k} = \mathsf{Align}_{\mathrm{native}}(x_{\hat{\sigma}_k}, \tilde{x}^{\mathrm{EDM}}_{0,k})$, and the native EDM update with step scale s_{step} continues from the projected endpoint:

$$x_{\sigma_{k+1}} = \bar{x}_{\hat{\sigma}_k} + s_{\mathrm{step}}(\sigma_{k+1} - \hat{\sigma}_k)\frac{\bar{x}_{\hat{\sigma}_k} - \tilde{x}^{\mathrm{EDM}}_{0,k}}{\hat{\sigma}_k}. \tag{29}$$

If the endpoint is unchanged, this is the original Boltz-2 sampling update.

C.2.1 LIGAND AND INTERMOLECULAR CONSTRAINTS

Ligand bond lengths and angles. The ligand's atom types, bond orders, and connectivity define its internal distance bounds. Boltz represents these through lower and upper bounds ℓ_{ij}, u_{ij} for bonded and 1–3 pairs. With relative padding η, the effective bounds and normalized residual are

$$\bar{\ell}_{ij} = (1-\eta)\ell_{ij}, \qquad \bar{u}_{ij} = (1+\eta)u_{ij},$$
$$r_{ij}^{\text{ligand}} = \left(\frac{[\bar{\ell}_{ij} - d_{ij}]_+}{\bar{\ell}_{ij}}, \frac{[d_{ij} - \bar{u}_{ij}]_+}{\bar{u}_{ij}} \right). \tag{30}$$

Intramolecular clashes use the lower-bound component with its own padding.

Ligand planarity and stereochemistry. Ring and double bond planarity use the fitted-plane residual in Eq. (8). Graph-ordered stereogenic tuples define the *E*/*Z* central-dihedral residual in Eq. (9), normalized by a fixed angular scale. Bond-tangent ChemGN (*R*/*S*) reduces the signed-margin residual in Eq. (11) with the bond-tangent correction in Eq. (13).

Protein–ligand exclusion. A ligand pose must leave sufficient space between nonbonded ligand and protein atoms. Let ρ_i, ρ_j be their van der Waals radii. For eligible protein–ligand pairs and relative padding η_{PL}, we use

$$d_{ij}^{\min} = (1-\eta_{\text{PL}})(\rho_i + \rho_j), \qquad r_{ij}^{\text{PL}}(x) = \frac{[d_{ij}^{\min} - \|x_i - x_j\|_2]_+}{d_{ij}^{\min}}. \tag{31}$$

Eligible pairs exclude covalently connected and single-atom chains and are limited to currently violated nearby pairs. Cartesian ChemGN reduces this exclusion residual by adjusting ligand coordinates while the protein is fixed. Rigid ligand $SE(3)$ projection reduces the same residual by translating and rotating the ligand as one body, preserving its internal geometry.

Protein-component separation. For complexes with multiple protein components, steric separation is imposed between components that can move independently. Covalently connected protein chains form one rigid component. Atom pairs in different components use Eq. (31) with the same padding. Components containing the same molecular entity also satisfy a centroid-separation requirement with minimum distance d_{cent}:

$$r_{PQ}^{\text{centroid}}(x) = \frac{[d_{\text{cent}} - \|\bar{x}_P - \bar{x}_Q\|_2]_+}{d_{\text{cent}}}. \tag{32}$$

The ligand-contacting receptor component is fixed, and the remaining eligible components undergo bounded $SE(3)$ projections. This reduces intercomponent overlap while preserving each component's internal coordinates.

C.2.2 PROJECTION SCHEDULE AND TERMINAL REFINEMENT

Projection order and schedule. Each denoising step projects the endpoint prediction immediately before the native update, in this order: the rigid-fragment *E*/*Z* projection, which restores the declared double bond configuration and refines its torsion; Cartesian ChemGN; the rigid-body ligand and protein-component $SE(3)$ projections; and bond-tangent ChemGN (*R*/*S*). The fraction of the sampling trajectory in which each projection acts is given in Table 7; the shared projections are defined in Appendix A.

Bounded corrections and acceptance. Each projection is bounded. Rigid-fragment refinement, rigid ligand $SE(3)$, and bond-tangent ChemGN (*R*/*S*) accept a correction only if it lowers their targeted residuals and passes checks on the other constraint families, and the protein-component projection limits any increase in protein–ligand exclusion. Exact settings are provided with the released code.

Terminal projection and confidence selection. After the last denoising step, a bounded terminal refinement applies the Cartesian ChemGN, protein-component, and bond-tangent ChemGN (*R*/*S*) projections without further denoising. Native confidence is recomputed on all five final complexes, and its maximum selects the top-1 pose.

Table 9: ***E/Z* accuracy by number of stereogenic double bonds on EZBench.** Best in **bold**, second-best underlined.

		All *E/Z* bonds $\leq 10^\circ$ (%) $\uparrow$			Median *E/Z* error ($^\circ$) $\downarrow$		
Number of *E/Z* bonds	N	Vanilla	Boltz-steering	AtomSteer	Vanilla	Boltz-steering	AtomSteer
One bond	813	64.6	66.2	**81.2**	5.63	5.53	**2.61**
Two bonds	192	53.7	48.4	**75.0**	6.46	6.96	**2.83**
Three or more	98	29.6	25.5	**70.4**	11.78	8.06	**2.19**

Comparison methods. Unsteered Boltz-2 and native-potential guidance each use one particle, five returned samples, and 200 steps. Boltz-steering uses native potentials with Feynman–Kac resampling over three interacting particles, as released (Singhal et al., 2025). Ishitani & Moriwaki applies restraint-based conjugate-gradient correction to the denoised ligand with 1,000 iterations (Ishitani & Moriwaki, 2025). The ProteinGS-based comparison applies 100 terminal Gauss–Seidel sweeps (Chen et al., 2026). All methods retain the checkpoint, seed, and native confidence selector and are evaluated on the same targets.

C.3 Tasks, data and additional results

EZBench. EZBench comprises 1,103 experimental protein–ligand complexes, each keyed by a PDB accession and a ligand identifier (Berman et al., 2000), selected so that every ligand carries at least one non-ring stereogenic double bond, and filtered by standard quality gates (resolution $\leq$ 2.5 Å, $R_{\text{free}} \leq 0.30$, ligand RSCC ≥ 0.8, RSR ≤ 0.25, occupancy ≥ 0.8, MW 100 to 1000 Da, ≥ 6 heavy atoms, non-covalent, drug-like, and a complete crystallographic model). For every retained ligand the reference *E/Z* configuration was read directly from the crystal geometry using RDKit's CIP labeler. The deposited coordinates are used as a post-generation reference. Since every target carries at least one evaluable stereogenic double bond, the *E/Z* rows use $N = 1{,}103$, and the *R/S* rows use the 538 targets with at least one evaluable stereocenter.

PoseBusters. Of the 308 complexes, 3 exhaust GPU memory for every method and 4 have no establishable crystallographic reference pocket for any method, giving 301 common targets. The *E/Z* and *R/S* subsets contain 21 and 177 targets. Release dates span 2021-01-13 to 2023-05-24, all before the Boltz-2 training cutoff.

Where the *E/Z* gain concentrates. Among the 1,103 EZBench ligands, 813 carry one stereogenic double bond, 192 carry two and 98 carry three or more, with 1,608 evaluable bonds in total. The fraction with every bond within 10°, for AtomSteer against Boltz-steering, is 81.2% against 66.2% on one-bond ligands, 75.0% against 48.4% on two-bond ligands, and 70.4% against 25.5% on ligands with three or more (Table 9). AtomSteer's margin *widens* with bond count, from 15.0 to 26.6 to 44.9 points. AtomSteer's median error is essentially flat across the three classes (2.61°, 2.83°, 2.19°) while Boltz-steering's degrades from 5.53° via 6.96° to 8.06°. The gain does not depend on the 10° threshold: AtomSteer leads at every threshold from 2.5° to 20° (Table 10), by 25.3 points at 2.5° narrowing to 11.2 at 20°, and the methods converge only past 30°, where basin assignment rather than angular precision is all that is being measured.

Temporal holdout. PDB release dates resolve for all 1,103 EZBench targets: 926 were released before the 2023-06-01 Boltz-2 training cutoff and 177 on or after it, and the *R/S* subset splits 451 and 87. Table 11 reports both splits, with metrics recomputed within each split. A release-date split does not establish sequence, pocket, or ligand novelty.

Qualitative examples. Figure 4 shows targets where AtomSteer repairs *E/Z* geometry and physical validity that Boltz-steering and the unsteered model leave incorrect, Figure 5 targets where the double bond is already close to correct and AtomSteer matches the baselines, and Figure 6 targets where Boltz-steering places the ligand closer to the crystal than AtomSteer.

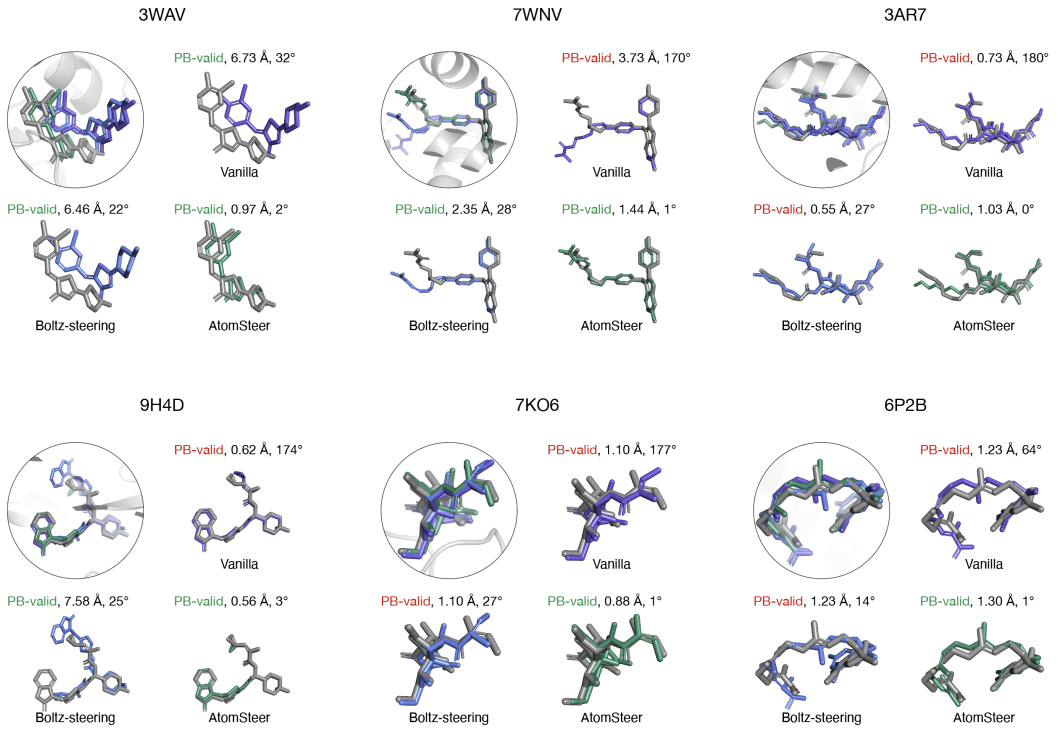


Figure 4: **Examples where AtomSteer resolves *E/Z* stereochemistry and pose validity that Boltz-steering and the unsteered model leave incorrect.** Each subpanel overlays the deposited crystal ligand (gray) with one model's confidence top-1 pose, placed in the crystal frame by aligning the protein; the PDB ID is shown at the top of each target and the circular inset shows all three methods together in the binding pocket. Above each pose are three annotations: PoseBusters validity (green = PB-valid, red = not PB-valid), ligand RMSD to the experimental ground truth (Å), and *E/Z* dihedral deviation of the double bond from the crystal structure (°). Methods: Vanilla (purple), Boltz-steering (blue), AtomSteer (green). Across all six targets AtomSteer reaches $\leq 3°$ *E/Z* deviation and a PB-valid pose: for 3AR7, 7KO6, and 6P2B the ligand is already well placed (RMSD ≤ 1.3 Å) but the double bond is mis-set by 14–180°, giving a twisted, PB-invalid geometry that AtomSteer repairs at essentially unchanged RMSD, while for 3WAV, 7WNV, and 9H4D, at least one baseline also misplaces the ligand (up to 7.6 Å) and AtomSteer recovers both a near-native pose (RMSD ≤ 1.4 Å) and correct stereochemistry.

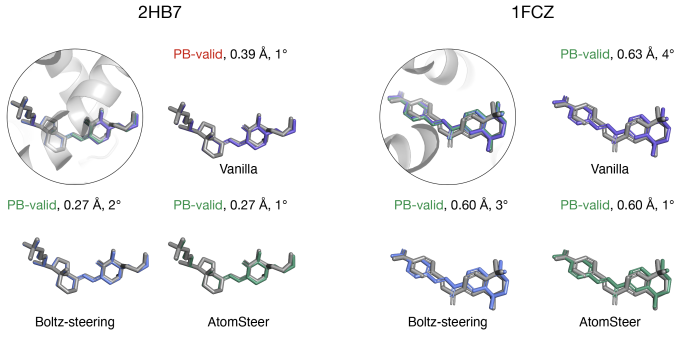


Figure 5: **Examples where the double bond is already well predicted and AtomSteer matches the baselines without degrading placement or validity.** Layout, colors, and annotations as in Figure 4. The *E/Z* geometry is close to correct in the unsteered model: all methods dock tightly (RMSD ≤ 0.6 Å) and are mostly PB-valid, and AtomSteer reduces the dihedral error slightly further.

Table 10: ***E/Z* threshold sensitivity on EZBench.** Percentage of the 1,103 ligands for which *every* evaluable stereogenic double bond falls within the given threshold of the deposited ligand, the criterion of Table 6 at 10°. Best per row in **bold**, second-best underlined.

Threshold ↑	Vanilla	Boltz-steering	**AtomSteer**
$\leq 2.5°$	21.7	21.8	**47.1**
$\leq 5°$	39.1	39.1	**63.4**
$\leq 10°$	59.6	59.5	**79.1**
$\leq 15°$	69.2	70.8	**85.9**
$\leq 20°$	73.3	78.6	**89.8**
$\leq 30°$	77.2	92.4	**93.0**
$\leq 45°$	79.6	**96.0**	95.6

Table 11: **EZBench split by PDB release date at the Boltz-2 training cutoff of 2023-06-01.** Same protocol, selector, and metric definitions as Table 6, recomputed within each split of the 1,103 targets. Best in **bold**, second-best underlined.

Split	Metric	Vanilla	Boltz-steering	**AtomSteer**
Before 2023-06-01 (N=926)	PB-valid ↑	51.2	90.6	**92.5**
	RMSD < 2 Å ↑	67.9	69.7	**69.9**
	Success@1 ↑	38.1	64.0	**65.1**
	E/Z median error (°) ↓	6.02	5.80	**2.48**
	All *E/Z* bonds $\leq 10°$ ↑	61.0	60.7	**80.5**
	Success@1+stereo ↑	27.8	39.8	**49.7**
On or after 2023-06-01 (N=177)	PB-valid ↑	50.3	97.2	**97.7**
	RMSD < 2 Å ↑	**38.4**	37.3	37.9
	Success@1 ↑	23.2	36.7	**37.3**
	E/Z median error (°) ↓	8.17	7.67	**3.22**
	All *E/Z* bonds $\leq 10°$ ↑	52.0	53.1	**72.3**
	Success@1+stereo ↑	13.0	19.2	**22.0**

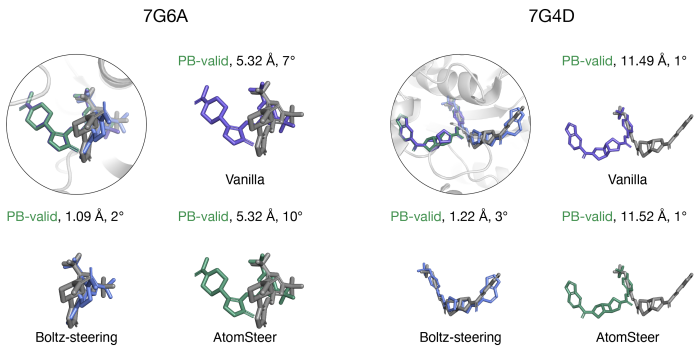


Figure 6: **Examples where Boltz-steering places the ligand closer to the crystal than AtomSteer.** Layout, colors, and annotations as in Figure 4. The double bond is within 10° in every method, and the AtomSteer pose is close to the Vanilla pose (ligand RMSD, Boltz-steering to AtomSteer: 7G4D 1.2 → 11.5 Å, 7G6A 1.1 → 5.3 Å).

Table 12: **OpenStructure evaluation of the same predictions on EZBench.** Computed with OpenStructure 2.12.0. The interface rows use the EZBench multi-chain subset. OpenStructure contains no stereochemical term, so these scores are a preservation check on protein and interface quality. Best in **bold**.

Metric	Vanilla	Boltz-steering	**AtomSteer**
lDDT-PLI ↑	0.73	**0.74**	0.73
Complex lDDT ↑	**0.95**	**0.95**	**0.95**
Backbone lDDT ↑	**0.98**	**0.98**	**0.98**
TM-score ↑	**0.98**	**0.98**	**0.98**
GDT-TS ↑	**0.97**	**0.97**	**0.97**
Cα RMSD (Å) ↓	**0.96**	**0.96**	0.97
Interface lDDT ↑	**0.77**	**0.77**	**0.77**
DockQ ↑	**0.75**	**0.75**	**0.75**

D EVALUATION PROTOCOLS AND METRICS

We collect the crystalline-material and molecular-conformer data, evaluation rules, and denominators here. Protein–ligand metrics and data are specified in Appendices C.1 and C.3.

D.1 CRYSTALLINE MATERIALS

D.1.1 MP-20 DE NOVO GENERATION

We generate 10,000 paired samples for vanilla FlowMM and AtomSteer from the same initial states and numbers of atoms.

Symmetry and relaxation metrics. We detect space groups in as-generated and UMA-relaxed endpoints with spglib (Togo et al., 2024) at $\mathrm{symprec} = 0.10$ Å and 5° angle tolerance. We define the *symmetry success rate* as the fraction of the full paired cohort assigned a space-group number of at least three, excluding $P1$ and $P\bar{1}$; an endpoint without a detected space group is a failure. The distribution distance d_{SG} is the base-2 Jensen–Shannon distance between the detected generated and MP-20 space-group distributions and therefore conditions on successful assignments; lower is better. Unique-SG counts likewise include detected groups only. An as-generated endpoint is geometrically valid (Valid) when its fractional coordinates and cell are finite, the cell is nonsingular, and its minimum periodic interatomic distance, including an atom's own periodic images, is at least 0.5 Å. Valid is evaluated before relaxation on all 10,000 samples, counts unevaluable endpoints as invalid, and includes no composition check. Forces and relaxations use the UMA-S-1p2 potential with its OMat task (Wood et al., 2025). For an endpoint x with UMA atomic forces $\mathbf{F}_i(x)$, $F_{\max}(x) = \max_i \|\mathbf{F}_i(x)\|_2$. The paired statistic $\Delta F_{\max}$ is the median over the 10,000 pairs of $F_{\max}(x^{\mathrm{AtomSteer}}) - F_{\max}(x^{\mathrm{FlowMM}})$, computed on as-generated endpoints before relaxation; FlowMM is the reference and has no entry. Full-cell relaxation starts from every as-generated endpoint and optimizes atomic positions and all cell parameters with the LBFGS optimizer, a Fréchet cell filter, force threshold 0.05 eV/Å, and a 100-step limit, without external pressure or symmetry constraints. Post-relaxation symmetry success rates evaluate all 10,000 relaxed endpoints per method regardless of optimizer convergence, which Appendix E reports.

D.1.2 MP-20 CRYSTAL STRUCTURE PREDICTION

The MP-20 test cohort contains 9,046 targets. Each locally run method generates 20 candidates per target. DiffCSP, CrystalFlow, and OMatG use their released code and MP-20 CSP checkpoints with their released sampling settings of 1,000 predictor–corrector steps, 100 Euler steps, and 210 Euler steps, respectively; source revisions are provided with the released code. Their saved outputs, our FlowMM run, and both AtomSteer variants are scored by the common validity-aware evaluator specified below. FlowMM and both AtomSteer variants start from the same saved initial noise for each target and candidate slot.

Structure matching and space-group recovery. The structure-match predicate follows the validity-aware StructureMatcher protocol of DiffCSP (Jiao et al., 2023) and FlowMM (Miller et al., 2024). A candidate matches its reference only when both pass the FlowMM validity check and the matcher returns a finite RMS distance. The validity check requires a minimum periodic interatomic distance of at least 0.5 Å, a cell volume of at least 0.1 Å^3, SMACT charge neutrality with the Pauling electronegativity test (single elements and all-metal compositions pass), and a successful CrystalNN local-environment fingerprint; crystal construction and preprocessing follow the released FlowMM evaluator. The matcher uses `stol=0.5`, `ltol=0.3`, `angle_tol=10`, primitive-cell reduction and scaling enabled, supercell matching and subset matching disabled, and a species comparator. Every planned target remains in the denominator; invalid, failed, or missing candidates are failures.

For a target i with saved candidates g_{ij},

$$\mathrm{MR@}K = \frac{1}{N}\sum_i \mathbf{1}[\exists j < K : \mathrm{match}(g_{ij}, r_i)]\,, \tag{33}$$

$$\mathrm{SG@}K = \frac{1}{N}\sum_i \mathbf{1}[\exists j < K : \mathrm{SG}(g_{ij}) = \mathrm{SG}(r_i)]\,, \tag{34}$$

$$\mathrm{Joint@}K = \frac{1}{N}\sum_i \mathbf{1}[\exists j < K : \mathrm{match}(g_{ij}, r_i) \wedge \mathrm{SG}(g_{ij}) = \mathrm{SG}(r_i)]\,. \tag{35}$$

Space groups are detected independently for candidate and reference at $\mathrm{symprec} = 0.10$ Å and 5° angle tolerance. Joint@K requires one candidate to satisfy both predicates. Our K=1 and K=20 results are nested prefixes of the same saved 20-sample pool; no reference-based ranking is used.

CSP baseline evaluation. We use a configuration-matched FlowMM run. For CDVAE and MCFlow, we report the published MR values. A dash marks a metric not evaluated under the common evaluator.

D.1.3 Motif-constrained generation

For each of Kagome, honeycomb, and triangular Archimedean lattices (ALs), we draw 1,000 conditions from the SCIGEN AL-specific atom-count and bond-length distributions and match them across vanilla FlowMM, AtomSteer, and SCIGEN. Vanilla FlowMM and AtomSteer also start from the same initial state for each condition, while SCIGEN shares the design conditions and runs its released code and MP-20 checkpoint with the unmodified 1,000-step sampler (source revision provided with the released code). The requested AL length is the larger of the released KDE draw and twice the selected element's metallic radius. All methods use the same full-cell UMA relaxation: UMA-S-1p2 with its OMat task, FIRE optimizer, Fréchet cell filter, force threshold 0.05 eV/Å, and 500-step limit, relaxing atomic positions and the full cell without external pressure, symmetry constraints, or post-relaxation reprojection. Relaxation covers every generated structure that can be built as a periodic crystal, independent of prescreening survival.

Constraint satisfaction, prescreening, and relaxation. Following SCIGEN (Okabe et al., 2026), we require the generated crystal to contain one of the three primary Archimedean lattices (ALs), triangular, honeycomb, or Kagome, with a specified magnetic element on the AL vertices at a specified bond length. AL constraint satisfied requires the requested magnetic element to form the requested AL at the prescribed nearest-neighbor distance in a geometrically valid crystal. A crystal is geometrically valid when it keeps the requested atom count, has atomic numbers from 1 to 100, a cell volume of at least 0.1 Å^3, and a minimum periodic interatomic distance, including self-images, of at least 0.5 Å. The detector searches same-species subsets under the 48 signed axis permutations, fits a common periodic translation, and accepts a Cartesian positional RMS of at most 0.15 Å, relative differences of at most 0.03 between the in-plane lengths a and b and between each length and the requested side length sd (Appendix B.1.2), and an in-plane angle within 2° of 120°. Following SCIGEN, prescreening survival is the fraction of generated structures that survive its released four-stage prescreening filter: charge neutrality, space occupancy ratio, a GNN classifier for energy above hull, and a GNN for stable structure distinction. Joint is the same-sample intersection of AL constraint satisfaction and prescreening survival. AL detected after relaxation is the fraction of conditions whose UMA relaxation converges to a geometrically valid endpoint containing the requested AL shape. The shape test applies the detector above without the requested-side-length condition and accepts any same-species subset, so it does not require the originally designated sites, raw AL constraint satisfaction, or prescreening survival. Every metric uses all 1,000 planned conditions per AL and method as denominator.

D.2 Molecular conformers

D.2.1 GEOM-Drugs test set and conformer budgets

Two molecules of the 1,000-molecule GEOM-Drugs test split have disconnected stored molecular graphs. Following the released ET-Flow evaluator, which skips disconnected graphs, we evaluate

the remaining 998 molecules, with 103,966 reference conformers and 207,932 reported conformers per method. The ODE50, SS50, and AtomSteer rows in Table 5 use these matched ensembles.

D.2.2 CONFORMER ACCURACY, VALIDITY, AND STEREOCHEMISTRY

Coverage and matching. For each GEOM-Drugs molecule, COV-R is the fraction of reference conformers within 0.75 Å symmetry-corrected heavy-atom RMSD of a generated conformer; COV-P reverses the reference and generated sets. AMR-R and AMR-P are the corresponding mean minimum RMSDs. We first compute each metric within a molecule and then summarize it by the mean and the median over all 998 molecules, so molecules contribute equally regardless of ensemble size. Table 5 reports means, and Table 13 reports both statistics.

Conformer validity. PoseBusters validity uses version 0.6.3 with the `mol` configuration (Buttenschoen et al., 2024). A conformer is valid only when all twelve checks pass: molecule loading, sanitization, InChI conversion, connectivity, absence of radicals, bond lengths, bond angles, internal steric clashes, aromatic-ring flatness, non-aromatic-ring non-flatness, double bond flatness, and internal energy. Unevaluable checks fail closed. For each molecule we evaluate the first $\min(2K, 32)$ generated conformers, where K is its number of reference conformers, and report the mean validity rate across molecules.

Stereochemical distribution error. For every molecule with an evaluable stereogenic double bond, we compute the one-dimensional Wasserstein distance between generated and held-out GEOM central-dihedral distributions in degrees, average over its evaluable bonds, and report the median across the 153 eligible molecules. The central dihedral of a bond is defined by the two double bond atoms and one substituent on each, and is taken as the unsigned angle in $[0^\circ, 180^\circ]$ between the two substituent bond vectors after projection perpendicular to the double bond, so 0° and 180° correspond to planar cis and trans placement of that substituent pair. A double bond is evaluable when RDKit assigns it an *E*/*Z* configuration after all hydrogens are removed and each of its two atoms carries a heavy-atom substituent, the same rule we apply to protein–ligand complexes (Appendix C.1). This excludes, for example, a terminal methylene whose two hydrogens are distinguished only by atom labels. This continuous W_1 endpoint captures distributional agreement without reducing each ensemble to a binary assignment.

D.2.3 GFN2-XTB PROTOCOL AND ENSEMBLE PROPERTIES

We evaluate 100 molecules sampled uniformly without replacement from the sorted GEOM-Drugs test set with NumPy `default_rng(42)`. For a molecule with K references, we use $\min(K, 32)$ reference conformers and $\min(2K, 32)$ generated conformers. GFN2-xTB 6.7.1 (Bannwarth et al., 2019) optimizes generated conformers with `--gfn 2 --opt normal` and then performs a single-point calculation with `--gfn 2 --iterations 1000 --dipole`; reference conformers receive the same single-point calculation without optimization. The molecular formal charge is passed with `--chrg`, and `--uhf` is set when required. All 100 molecules have complete ensembles for every method in Table 5.

GFN2-xTB ensemble properties. We evaluate the four GFN2-xTB ensemble properties used by ET-Flow (Hassan et al., 2024): Boltzmann-weighted mean energy E, dipole moment μ, HOMO–LUMO gap $\Delta\epsilon$, and minimum energy $E_{\min}$. We compare each generated ensemble with its GEOM reference ensemble and report the median molecular absolute error. The Boltzmann weights use $kT = 0.592$ kcal/mol.

E ADDITIONAL CRYSTALLINE-MATERIALS RESULTS

Sampling and projection details are in Appendix B.1. Evaluation protocols and denominators are in Appendix D.1.

E.1 De novo generation: size, energy, and relaxation

Figure 2 groups structures by the number of atoms in the generated unit cell, which is fixed before sampling and shared within each paired sample. UMA single-point force comparisons before relaxation pair all 10,000 generated samples; AtomSteer lowers $F_{\max}$ in 74.5% of pairs, with a median change of -0.0394 eV/Å. Energy comparisons use the 9,984 same-composition pairs; their median change is -2.26×10^{-4} eV/atom. Full-cell relaxation converges for 80.48% of FlowMM and 82.16% of AtomSteer endpoints.

E.2 Crystal structure prediction: size analysis

Figure 7 reports target-space-group recovery by crystal size.

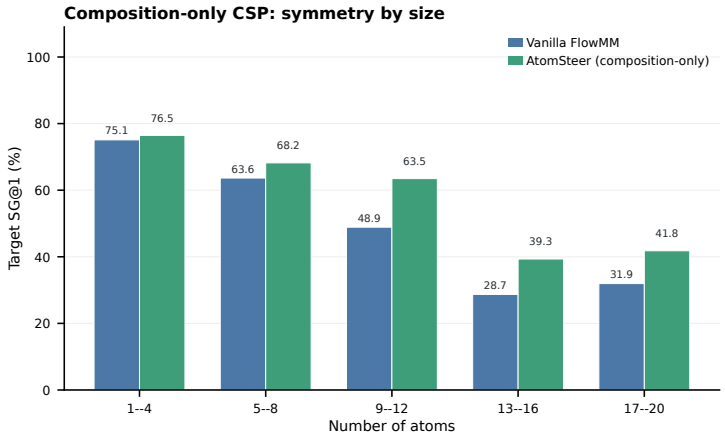


Figure 7: **AtomSteer (composition-only) improves target-space-group recovery across MP-20 crystal sizes.** Target SG@1 is evaluated on all 9,046 targets and grouped by the number of atoms in the generated unit cell. AtomSteer infers the symmetry of both L and F from FlowMM's endpoint prediction and projects both.

E.3 Motif-constrained generation examples

Table 4 compares AtomSteer with SCIGEN, while Figure 8 visualizes matched vanilla FlowMM and AtomSteer structures together with the requested AL and the relaxed AtomSteer endpoint.

F Additional molecular-conformer results

Sampling and projection details are in Appendix B.2. Evaluation protocols and denominators are in Appendix D.2.

F.1 Mean and median coverage and matching

Table 13 reports the mean and median of each coverage and matching metric over all 998 molecules.

Table 13: **Mean and median coverage and matching on all 998 GEOM-Drugs test molecules.** Each metric is computed within a molecule and summarized over molecules. COV uses a 0.75 Å threshold. Best results are **bold** and second-best results are underlined.

	COV-R (%) ↑		AMR-R (Å) ↓		COV-P (%) ↑		AMR-P (Å) ↓	
Method	Mean	Median	Mean	Median	Mean	Median	Mean	Median
ET-Flow ODE50	79.92	84.52	0.458	0.436	73.40	80.31	0.562	0.500
ET-Flow SS50	**79.93**	**84.57**	**0.456**	**0.435**	73.52	80.19	0.559	0.497
AtomSteer	79.01	82.90	0.457	0.440	**75.49**	**82.65**	**0.529**	**0.463**

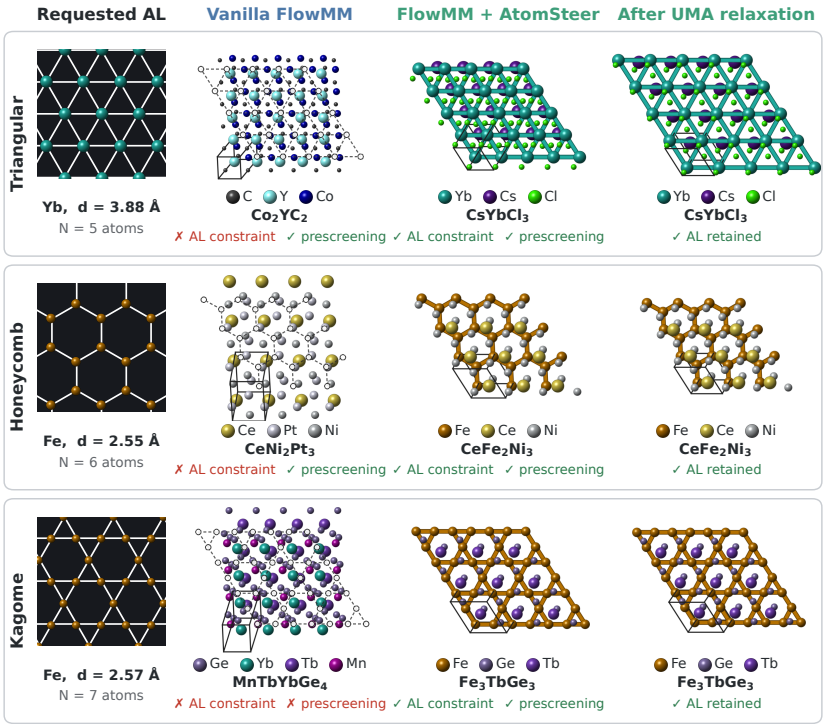


Figure 8: **Full motif-constrained generation examples.** Each row fixes the target element, nearest-neighbor distance d, and atom count N, then shows the requested AL, matched vanilla FlowMM and AtomSteer endpoints, and the AtomSteer structure after UMA relaxation. Structures use a shared tilt and length scale within each row; bonds connect AL vertices, and the wireframe marks the unit cell.

F.2 Qualitative examples

Figure 9 shows test molecules on which an ET-Flow conformer violates a graph constraint that AtomSteer satisfies, and Figure 10 shows AtomSteer conformer ensembles against the GEOM reference.

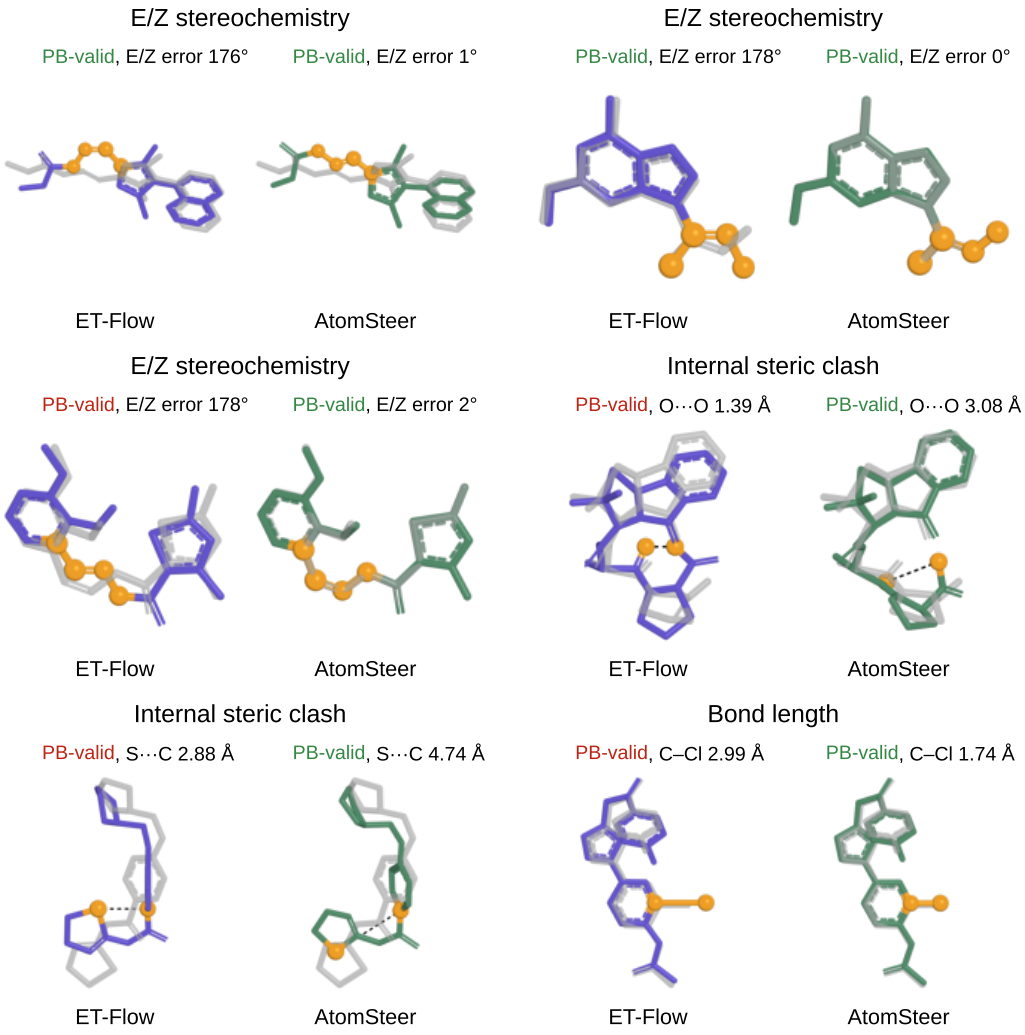


Figure 9: **Representative GEOM-Drugs test molecules on which AtomSteer satisfies a constraint that ET-Flow violates.** Each pair shows an ET-Flow ODE50 conformer (purple) and an AtomSteer conformer (green) of the same molecule, both aligned to the same GEOM reference conformer (gray); highlighted atoms mark the constrained double bond, atom pair, or bond. Above each conformer are its PoseBusters validity (green = PB-valid, red = not PB-valid) and its constraint error: the *E*/*Z* dihedral deviation from the target configuration, the nonbonded distance, or the bond length. PoseBusters does not test *E*/*Z* configuration, so an *E*/*Z* violation can pass it. Each AtomSteer conformer is the PB-valid conformer of that molecule closest in heavy-atom RMSD to the ET-Flow conformer shown.

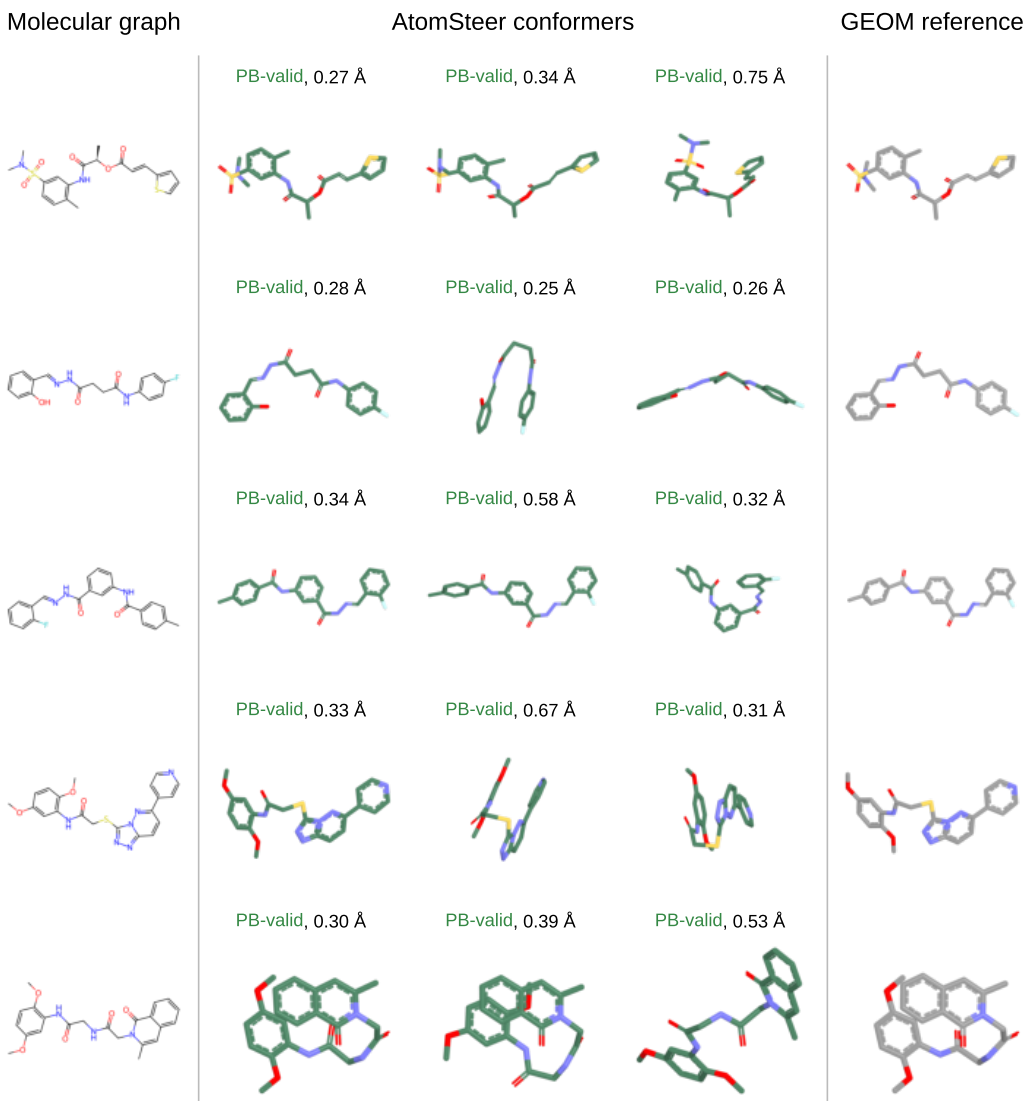


Figure 10: **AtomSteer conformer ensembles for five GEOM-Drugs test molecules.** Each row shows the molecular graph, three AtomSteer conformers, and the GEOM reference conformer nearest to the first. Above each conformer are its PoseBusters validity and its symmetry-corrected heavy-atom RMSD to the nearest reference conformer. The shown conformers are PB-valid, are mutually distinct, and lie 0.25–0.75 Å from their nearest reference.