TIME-DEPENDENT DIRECTIONAL GUIDANCE FOR PHYSICS-INFORMED MOLECULAR DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have become effective generators of 3D molecular structures, yet geometric plausibility alone does not ensure physical suitability for structure-based design. A common remedy is to add energy-based guidance during sampling. We argue that this naive intervention can behave like a forceful external correction rather than a compatible part of the reverse diffusion process, producing two critical failure modes: temporal misalignment, where a fixed physical signal ignores the coarse-to-fine denoising rhythm, and directional conflict, where raw energy gradients oppose chemically plausible generative motion. We propose TDGM, a Time-Dependent Directional Guidance Model that provides a plug-and-play, lightweight framework for aligning physical guidance with both denoising time and update direction. TDGM uses a temporal schedule to synchronize guidance strength with different generation stages, a directional projection to filter conflicting gradient components, and multi-scale physical priors to cover geometry, steric safety, pocket interaction, and global placement. On 100 CrossDock pockets with 10,000 generated molecules, TDGM obtains the best Vina docking score, QED, and Lipinski score among the compared methods, while MolPilot remains strongest on diversity and normalized SA. The results show that properly scheduled and directionally filtered physical guidance improves binding-oriented and drug-likeness endpoints while preserving explicit trade-offs across molecular statistics.

1 INTRODUCTION

Diffusion models have become powerful generative models of 3D molecular structures because they learn to reverse stochastic noising over coordinates and molecular graphs (Qin et al., 2025a; Zhang et al., 2025; Liu et al., 2025a; Jung et al., 2024; Chen et al., 2025a; Fan et al., 2024). Yet structure-based design demands more than merely plausible geometry. A ligand may appear realistic while violating ideal bond geometry, forming suboptimal contacts with the binding pocket, or exhibiting pronounced strain after energy minimization.

Physics-aware molecular generation addresses this gap through geometry-aware architectures, property guidance, reinforcement learning, and energy-based sampling (Zhang et al., 2024; Zeng et al., 2026; Oestreich et al., 2025). Yet sampling-time energy guidance often assumes that a physical signal can simply be attached and strengthened. Direct injection can conflict with learned reverse dynamics, improving one endpoint while degrading another or failing practical filters (Sahoo & Hanson, 2025; Jian et al., 2026a). This limitation can be traced to two critical failure modes. *Temporal misalignment* occurs when a fixed guidance strength ignores that early reverse steps form coarse structure while later steps refine local chemistry. *Directional conflict* occurs when raw energy gradients improve a target score but point against chemically plausible diffusion directions. Thus the key questions are when physical guidance should act and which gradient components should modify the sample.

To address these challenges, we propose **TDGM** (Time-Dependent Directional Guidance Model), a lightweight test-time framework for diffusion-based molecular generation in protein pockets. TDGM addresses three limitations of naive physics-guided sampling. First, it leverages time-

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.

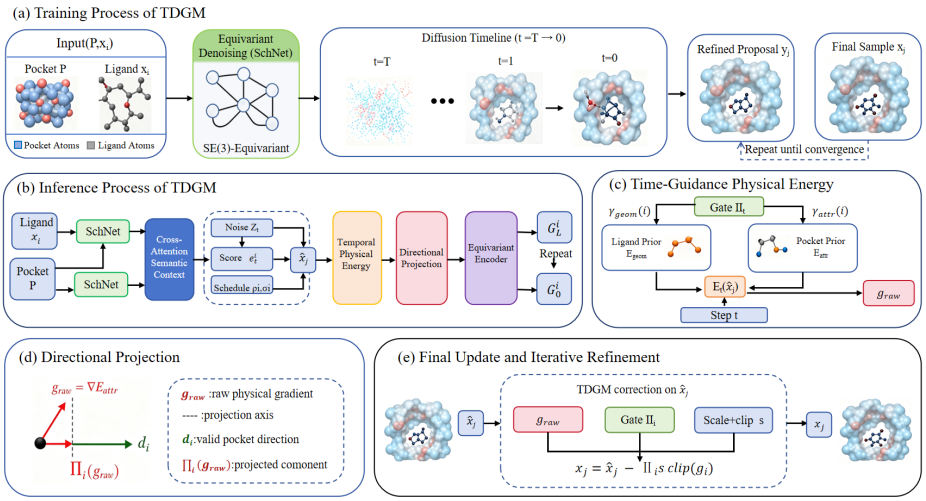

Figure 1: TDGM sampling framework. (a) The base SE(3)-equivariant diffusion sampler maps the pocket-conditioned state (P, x_i) along the reverse diffusion timeline. (b) TDGM inserts a guidance kernel between the denoising proposal and the accepted next sample. (c) The temporal physical energy combines geometry and pocket priors with step-dependent gates and adaptive weights. (d) Directional projection filters raw physical gradients through valid pocket-interaction directions. (e) The refined proposal is passed through equivariant encoding and iterative refinement until the final sample is obtained.

dependent guidance scheduling to reduce the temporal mismatch between physical energy and reverse diffusion, preserving early scaffold formation while strengthening local refinement at later denoising stages. Second, it introduces directional projection to alleviate conflicts between raw energy gradients and chemically plausible generative updates, retaining pocket-compatible components while suppressing conflicting directions. Third, it integrates a set of multi-scale physical priors spanning bond geometry, ligand and protein–ligand sterics, pocket attraction, and global centering, thereby characterizing molecular feasibility across complementary physical scales. Together, these components convert physical guidance from an unconstrained external force into a scheduled, directionally aligned, and pocket-aware correction. Figure 1 illustrates how TDGM inserts this guidance kernel into the base diffusion sampler without retraining the generator. Our key contributions are as follows:

- We formulate physics-guided molecular diffusion as an alignment problem and identify temporal misalignment and directional conflict as two major failure modes of naive energy guidance.
- We propose TDGM, a plug-and-play test-time framework that combines temporal scheduling, directional projection, and multi-scale physical energy to guide PMDM-based molecular generation without retraining the base sampler.
- We validate TDGM using 100 CrossDocked2020-derived protein pockets, demonstrating that the proposed guidance strategy improves binding-oriented and drug-likeness outcomes over the PMDM base sampler while controlling molecular-size effects.

2 Related Work

Related work is reviewed along three aspects central to TDGM: pocket-conditioned generative models, molecular feasibility constraints, and training-free sampling guidance, corresponding respectively to the base generator, physical objectives, and inference-time guidance mechanism.

Pocket-Conditioned Generative Models. Diffusion and flow models represent coordinates, molecular graphs, and ligand–pocket complexes jointly (Schneuing et al., 2024; Huang et al., 2024; Qin et al., 2025a; Zhang et al., 2025; Liu et al., 2025a), while equivariant architectures respect 3D molec-

ular symmetries (Schutt et al., 2018; Gasteiger et al., 2022). These backbones provide strong generative priors, but samples may still violate geometry, pocket compatibility, or medicinal-chemistry filters.

Multi-Objective Molecular Constraints. Generated ligands must balance affinity, geometry, sterics, drug-likeness, synthesizability, and diversity. Existing methods use topology, pharmacophores, property or affinity guidance, force-field terms, and energy regularization (Hu et al., 2025; Sako et al., 2026; Han et al., 2024; Cremer et al., 2024; Oestreich et al., 2025; Jian et al., 2026b). Because improving one objective can affect others (Qin et al., 2025b; Liu et al., 2025b; Chen et al., 2025b), evaluation should report joint feasibility rather than a single score.

Training-Free Sampling Guidance. Classifier(-free), universal, and gradient guidance steer pretrained diffusion models without retraining (Dhariwal & Nichol, 2021; Ho & Salimans, 2022; Bansal et al., 2024; Guo et al., 2024). Molecular variants support property control, target-aware generation, and multi-objective design (Han et al., 2024; Hu et al., 2025; Chen et al., 2025b), but leave open how physical gradients should be scheduled and filtered in pocket-conditioned diffusion.

3 METHOD

Given a target pocket P, TDGM refines a noisy ligand state into a pocket-conditioned molecular sample. It is applied as a test-time guidance layer on top of the pretrained PMDM sampler (Huang et al., 2024). As shown in Figure 1, PMDM first predicts a denoising proposal, and TDGM refines it using step-dependent physical priors and directionally filtered gradients before the next state is accepted.

At reverse step i, the current ligand state is x_i. PMDM produces a raw proposal $\hat{x}_j$, TDGM converts it into a refined proposal y_j, and the equivariant update block accepts the next state x_j. This notation separates the learned proposal, the guidance-corrected proposal, and the accepted sample, specifying where TDGM is inserted in the reverse diffusion trajectory.

3.1 POCKET-CONDITIONED SAMPLING

The base sampler first produces an intermediate denoising proposal for subsequent guidance. PMDM follows an SE(3)-equivariant reverse diffusion process that predicts both local atom-level refinement and global pocket-conditioned motion from the current ligand state and pocket structure. The local field updates fine geometry, while the global field captures pocket-level displacement. TDGM preserves the learned denoising step and applies physical correction to the resulting proposal.

Denote by $x_i \in \mathbb{R}^{N\times 3}$ the ligand coordinates at reverse diffusion index i, and by $P \in \mathbb{R}^{M\times 3}$ the centered protein pocket coordinates. The base sampler combines local atom-level refinement and global pocket-conditioned motion as

$$\Delta x_i = w_{\mathrm{loc}} u_{\mathrm{loc}}(x_i, P, i) + w_{\mathrm{glob}} u_{\mathrm{glob}}(x_i, P, i), \tag{1}$$

where w_{loc} and w_{glob} specify the relative weights of the local and global coordinate updates, respectively.

The base reverse step then constructs a temporary next-state proposal

$$\hat{x}_j = x_i + \rho_i \Delta x_i + \sigma_i z_i, \quad j < i, \tag{2}$$

where ρ_i specifies the weight of the learned denoising direction, $\sigma_i z_i$ denotes the stochastic diffusion component, and both coefficients are determined by the fixed reverse schedule. This proposal retains the learned PMDM dynamics while exposing a single interface for TDGM to refine the sample before acceptance.

3.2 TIME-GATED GUIDANCE

The effect of physical guidance varies across reverse diffusion steps. Early steps mainly form the coarse scaffold, whereas later steps refine clashes, contacts, and pocket placement. TDGM therefore

introduces a temporal gate to activate physical correction at step i before evaluating the corresponding energy terms.

We define a binary temporal gate

$$I_i = \mathbf{1}\{i_{\min} \le i \le i_{\max}\}, \tag{3}$$

where $i_{\min}$ and $i_{\max}$ specify the active guidance interval. The temporal gate reflects a modeling assumption that physical guidance contributes most effectively at selected reverse steps, rather than uniformly throughout the reverse process. Outside this interval, the sampler follows the base denoising dynamics without physical correction.

For the main physical-guidance setting, TDGM evaluates differentiable physical terms on the proposal $\hat{x}_j$. For one generated ligand, $\mathcal{B}$ denotes the ligand bond set, $R_P = \mathrm{mean}_m \|p_m - \bar{p}\|_2$ denotes the pocket radius, and $\bar{x}$ and $\bar{p}$ denote the ligand and pocket centroids. The guidance terms cover bond-length regularization, intra-ligand clash suppression, protein–ligand clash suppression, pocket-attraction, and centroid alignment.

These terms are not intended as a complete molecular force field; instead, they impose complementary differentiable constraints. The bond term regularizes bond-length deviations, the intra-ligand and protein–ligand clash terms suppress internal and interfacial overlaps, the pocket-attraction term promotes pocket-compatible placement, and the centering term limits drift from the binding site. Together, these priors steer the proposal while retaining compatibility with the learned diffusion prior.

3.3 Multi-Scale Energy

After guidance activation, TDGM assigns step-dependent weights to the physical priors, reflecting the different constraints required for coarse scaffold formation and late local refinement. Guidance is weak during global scaffold formation, increases during topology and contact refinement, and becomes stronger for late local geometry and pocket placement. The temporal physical energy contains two groups: the geometry prior models local chemical feasibility through bond consistency and clash avoidance, while the pocket prior models binding-site compatibility through pocket attraction and global placement. In the four-term setting, these two groups are

$$\begin{aligned} E_{\mathrm{geom}}(i) = \gamma_{\mathrm{geom}}(i)\big(&\lambda_{\mathrm{bond}} E_{\mathrm{bond}} + \lambda_{\mathrm{lig}} E_{\mathrm{lig}} \\ &+ \lambda_{\mathrm{pl}} E_{\mathrm{pl}}\big), \end{aligned} \tag{4}$$

$$\begin{aligned} E_{\mathrm{attr}}(i) = \gamma_{\mathrm{attr}}(i)\big(&\lambda_{\mathrm{pocket}} E_{\mathrm{pocket}} \\ &+ \lambda_{\mathrm{center}} E_{\mathrm{center}}\big). \end{aligned} \tag{5}$$

The λ coefficients define the relative importance of individual priors, while $\gamma_{\mathrm{geom}}(i)$ and $\gamma_{\mathrm{attr}}(i)$ determine how strongly geometry and pocket attraction are emphasized at step i. The active physical energy is then

$$E_i(\hat{x}_j, P) = I_i \left(E_{\mathrm{geom}}(i) + E_{\mathrm{attr}}(i)\right). \tag{6}$$

Thus, the physical objective is coupled to denoising time rather than imposed uniformly. This scheduling is the mechanism by which TDGM aligns physical correction with the coarse-to-fine behavior of reverse diffusion.

For the alternative three-stage guidance family, the trajectory is divided into placement, contact, and clash-refinement stages:

$$\begin{aligned} E_{\mathrm{3stage}}(i) = &\lambda_{\mathrm{place}} w_{\mathrm{place}}(i) E_{\mathrm{place}} \\ &+ \lambda_{\mathrm{contact}} w_{\mathrm{contact}}(i) E_{\mathrm{contact}} \\ &+ \lambda_{\mathrm{clash}} w_{\mathrm{clash}}(i) \\ &\quad E_{\mathrm{clash}}. \end{aligned} \tag{7}$$

This variant follows the same motivation: early guidance emphasizes coarse placement, middle guidance encourages pocket contact formation, and late guidance removes local clashes. The main experiments focus on the four-term projected setting, while ablations test schedule variants.

3.4 DIRECTIONAL PROJECTION

Temporal scheduling controls when guidance acts, but an energy-reducing gradient may still be incompatible with the selected pocket-contact geometry. TDGM therefore filters the physical gradient before applying it to the proposal.

The raw physical gradient of the active physical energy E_i at step i is given by

$$g_{\text{raw}} = \nabla_{\hat{x}_j} E_i(\hat{x}_j, P). \tag{8}$$

This gradient gives the steepest physical-energy descent direction, but only part of it may be useful for valid molecular refinement. Components that are unrelated to feasible pocket interactions can introduce unstable or chemically implausible corrections.

To filter these components, TDGM constructs a pocket-aware direction d_i for each ligand atom:

$$d_i = \text{norm}\left(d_{\text{pocket}}(\hat{x}_j, P)\right). \tag{9}$$

The direction d_{pocket} represents locally valid protein–ligand interaction directions. It turns the physical update from unconstrained energy descent into a pocket-aware correction that favors feasible contacts.

The physical gradient is then projected as

$$\Pi_i(g_{\text{raw}}) = \langle g_{\text{raw}}, d_i \rangle d_i. \tag{10}$$

This projection preserves the component along the chosen pocket-aware direction and removes its orthogonal component; it does not by itself guarantee alignment with the learned PMDM update. After optional norm clipping, the refined proposal is

$$y_j = \hat{x}_j - s\,\text{clip}\left(\Pi_i(g_{\text{raw}})\right), \tag{11}$$

where s is the guidance scale. The base model proposes $\hat{x}_j$, whereas TDGM converts it into a temporally scheduled, pocket-direction-filtered proposal y_j.

3.5 EQUIVARIANT REFINEMENT

After physical correction, the resulting state y_j is incorporated into the equivariant update path. This preserves the PMDM sampling structure while allowing TDGM to repeat a proposal–guidance–acceptance cycle throughout the reverse trajectory.

The refined proposal y_j is passed through the SE(3)-equivariant update block, and the accepted next state is written as

$$x_j = R_\theta(y_j, P, i), \tag{12}$$

where R_θ denotes the equivariant refinement/update operation. Repeating this procedure from $t = T$ to $t = 0$ yields the iterative refinement process in Fig. 1e. Thus each proposal is filtered by time-dependent physical priors and chosen pocket-interaction directions before acceptance.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

The evaluation assesses TDGM from three perspectives: aggregate performance, component necessity, and mechanistic behavior. CrossDocked2020-derived pockets (Francoeur et al., 2020) evaluate structure-based generation at scale, comparisons with PMDM isolate test-time guidance under the same base sampler, and case plus trajectory analyses connect quantitative gains to interpretable interactions and staged denoising.

Baselines and protocol. TDGM is evaluated on 100 protein pockets with 100 generated molecules per pocket. The base generator is fixed to PMDM, and TDGM adds guidance only at sampling time without retraining. We compare against AR-SBDD (Luo et al., 2021), EDM (Hoogeboom et al., 2022), DiffSBDD (Schneuing et al., 2024), PMDM (Huang et al., 2024), DiffPharma (Sako et al., 2026), and MolPilot (Qiu et al., 2025). Internal variants are diagnostic only, and QM9 (Ramakrishnan et al., 2014) is reserved for exploratory schedule analysis.

Table 1: Main CrossDock comparison on 100 pockets and 10,000 generated molecules. Vina is better when lower; up arrows indicate larger is better. Lip., Div., and MW denote Lipinski, diversity, and molecular weight. Missing values were not reported by the corresponding baseline.

Method	Vina	QED ↑	SA ↑	LogP	Lip. ↑	Div. ↑	MW
AR-SBDD	-6.215	0.502	0.421	0.257	4.795	0.679	331.98
EDM	-6.351	0.514	0.455	0.351	4.639	0.574	–
DiffSBDD	-6.584	0.495	0.342	–	4.795	0.730	–
PMDM	-7.574	0.524	0.557	0.410	4.862	0.754	324.38
DiffPharma	-7.057	0.502	0.662	0.737	4.715	0.784	265.16
MolPilot	-7.102	0.524	**0.720**	**1.193**	4.746	**0.906**	295.03
TDGM	**-7.628**	**0.673**	0.562	1.082	**4.951**	0.667	285.34

Metrics and evidence. The metrics cover binding-oriented placement (Vina), drug-likeness (QED and Lipinski), synthetic accessibility (normalized SA), physicochemical distribution (LogP), distributional spread (diversity), and size-driven docking effects (molecular weight). The 100-pocket comparison is the primary aggregate evaluation. The 200-molecule ablation is exploratory, the four docking cases are qualitative examples, and the single-pocket trajectory is a descriptive mechanism diagnostic. These evidence levels are reported separately so that case-level behavior is not treated as population-level validation.

4.2 Main Results

Table 1 reports binding-oriented, drug-likeness, synthesis-proxy, diversity, and size metrics on 100 pockets. TDGM has the best Vina (-7.628), QED (0.673), and Lipinski score (4.951) among the listed methods, whereas MolPilot has higher normalized SA and diversity. The improvement is therefore not uniform across molecular objectives.

PMDM is the shared-sampler counterpart. Relative to PMDM, TDGM lowers Vina from -7.574 to -7.628 while decreasing mean molecular weight from 324.38 to 285.34, arguing against a gain explained solely by larger ligands. This comparison supports the complete TDGM guidance package, but does not isolate the projection component.

4.3 Ablation Study

The 200-molecule ablation examines the contribution of complete and controlled physical guidance across unguided sampling, single-term correction, atom-prior guidance, and complete-guidance variants. Table 2 suggests that removing guidance leads to weak binding and chemical quality. Early clash-only correction yields limited gains: it reduces severe steric overlap, but does not provide the topology- and pocket-aware constraints needed for plausible ligand formation within the binding site.

Variants with more structured physical priors perform better than no-guidance and clash-only sampling. Specifically, no guidance yields Vina -6.166, QED 0.302, and Q+V 1.0%, whereas atom-prior guidance reaches -6.805, 0.628, and 29.0%, respectively. The complete variants remain close: full early, full late, and full uniform obtain Vina scores of -6.860, -6.821, and -6.888, with Q+V rates of 32.0%, 31.0%, and 34.5%. Their 100.0% pocket values denote threshold-based geometric feasibility rather than binding strength or specificity, and the small differences preclude a universal scheduling claim from this 200-molecule ablation.

4.4 Case Studies

The case analysis provides qualitative evidence that the favorable aggregate scores correspond to interpretable ligand–pocket interactions. Figure 2 and Table 3 are interpreted together, as the figure presents generated scaffolds and docked environments and the table reports QED, Vina, Bond MSE, and residue-level contacts. TDGM-096-3 targets the HLA-B*57:01 peptide-binding groove and

Table 2: Exploratory 200-molecule schedule ablation. Q+V is the fraction of molecules satisfying both QED and docking thresholds.

Variant	Vina	QED	SA	Q+V	Pocket
No guidance	-6.166	0.302	0.440	1.0	40.0
Clash early	-6.205	0.300	0.441	1.5	40.0
Atom prior	-6.805	0.628	0.516	29.0	100.0
Full early	-6.860	0.621	0.509	32.0	100.0
Full late	-6.821	0.625	0.519	31.0	100.0
Full uniform	**-6.888**	0.624	0.506	**34.5**	100.0

Table 3: Representative docking cases. QED measures drug-likeness, B-MSE is bond-geometry error, and C/H denotes local contacts/hydrogen bonds.

Compound	QED	Vina	B-MSE	C/H
TDGM-096-3	0.83	-9.66	0.029	28/1
TDGM-091-9	0.77	-9.36	0.005	26/0
TDGM-095-1	0.83	-8.66	0.008	26/3
TDGM-093-9	0.85	-8.72	0.008	19/1

shows the strongest displayed docking score, with high QED, controlled bond error, dense local contacts, and one hydrogen bond.

The remaining cases illustrate distinct interaction profiles. TDGM-095-1 combines high QED, low Bond MSE, and three predicted hydrogen bonds, whereas TDGM-091-9 and TDGM-093-9 show favorable Vina scores, low geometric errors, and multiple predicted contacts. These examples show that favorable aggregate metrics can coexist within selected molecules, but do not independently establish the guidance mechanism or experimental binding efficacy.

4.5 Temporal Trajectory Analysis

Figure 3 depicts the reverse-denoising trajectories of TDGM and PMDM on the same pocket, with columns ordered from the final ligand state to the initial coarse state (100% to 0%). The reported RMSD and Bond MAE quantify deviations of each intermediate frame from the corresponding model's final generated ligand; therefore, the figure characterizes trajectory organization rather than absolute experimental pose accuracy. Clear differences appear at the early 0% and 30% stages. At 0%, TDGM preserves comparable global proximity to PMDM (RMSD 1.18 vs. 1.21 Å) while reducing local bond distortion (Bond MAE 0.52 vs. 0.66 Å). At 30%, the two methods remain similar in RMSD (0.71 vs. 0.70 Å), whereas TDGM maintains lower Bond MAE (0.41 vs. 0.50 Å). These results support the temporal-scheduling rationale that early guidance should favor coarse pocket-compatible placement while avoiding excessive internal geometric distortion.

The transition from 30% to 70% further indicates a coarse-to-fine refinement pattern. TDGM reduces Bond MAE from 0.41 to 0.18 and then to 0.06 Å, suggesting progressively stronger local geometric repair after the initial scaffold has formed. Consistently, the generated ligand evolves from a rough contact-forming configuration toward a more stable final orientation, matching the intended phase allocation of early placement, intermediate topology and bond refinement, and late pose consolidation. This interpretation remains case-specific rather than absolute, since PMDM attains slightly lower RMSD and Bond MAE at the 50% frame. The stronger conclusion is that TDGM produces a more phase-specific and geometry-preserving trajectory during coarse generation and contact refinement, while both samplers ultimately converge to their own final ligand states. The terminal 100% RMSD and Bond MAE values are zero by construction because each final generated frame serves as its own trajectory reference.

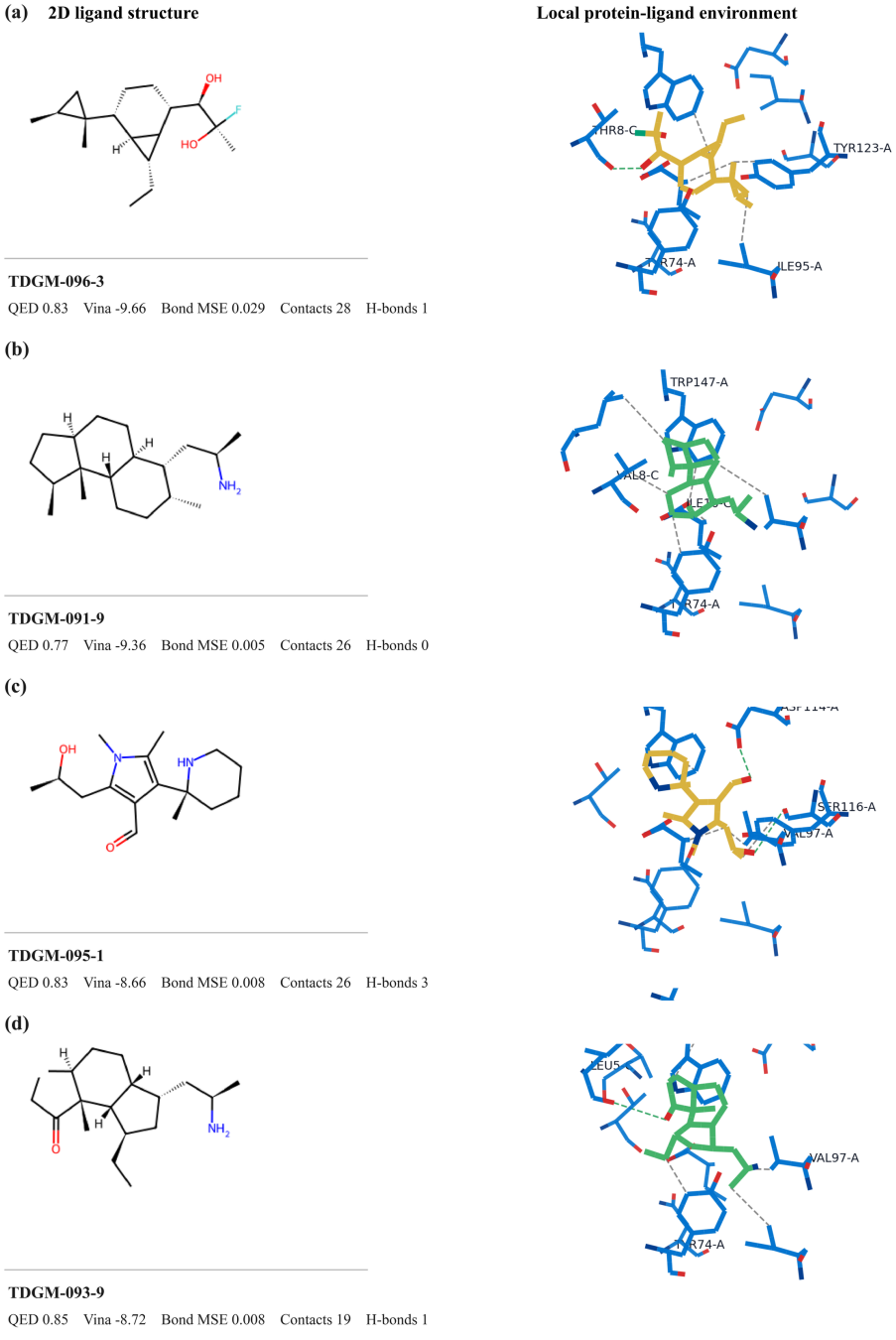

Figure 2: Representative TDGM docking cases. The left panel preserves the full 2D ligand structure; the right panel shows the local protein–ligand environment. Yellow/green sticks are ligands, blue sticks are interacting pocket residues, and dashed lines are predicted contacts or hydrogen bonds. The four rows are read together with Table 3: they test whether favorable docking scores coexist with valid chemistry and explicit local interactions.

4.6 Discussion

The complete TDGM package improves Vina from -7.574 to -7.628, QED from 0.524 to 0.673, and Lipinski score from 4.862 to 4.951 relative to PMDM, while reducing mean molecular weight

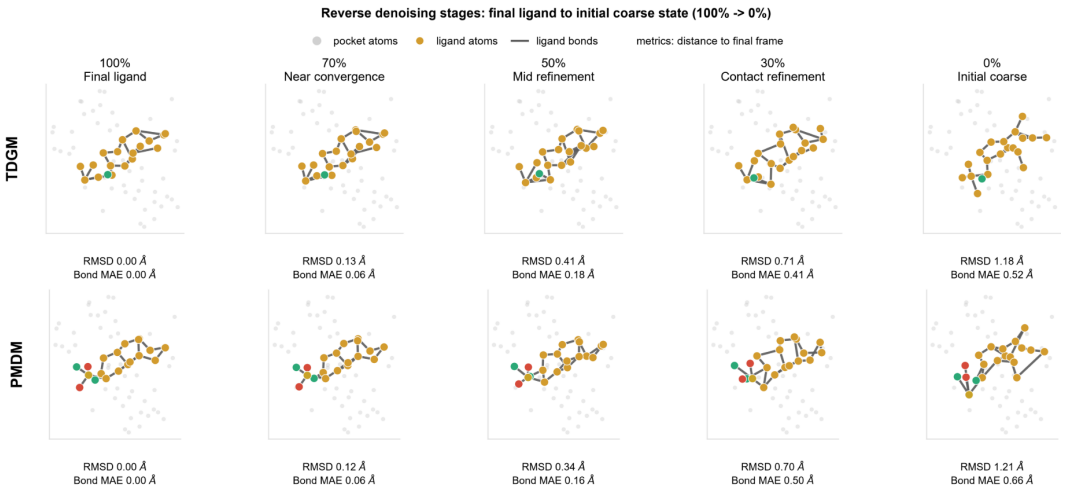

Figure 3: TDGM and PMDM denoising trajectories on the same pocket. Progress follows reverse diffusion from $t = T$ (0%) to $t = 0$ (100%); RMSD and Bond MAE are measured against each method's own final generated ligand.

Table 4: Temporal trajectory metrics for TDGM-096-3. Prog. runs from $t = T$ to $t = 0$; B-MAE, Cont., and PL-d denote bond MAE, contact score, and protein–ligand distance.

Stage	Prog.	RMSD	B-MAE	Cont.	PL-d
Early	0%	1.18	0.52	1.00	2.38
Early	30%	0.71	0.41	0.90	1.52
Middle	50%	0.41	0.18	0.95	1.74
Late	70%	0.13	0.06	0.90	1.81
Late	100%	0.00	0.00	0.90	1.80

from 324.38 to 285.34. However, normalized SA changes only from 0.557 to 0.562 and diversity decreases from 0.754 to 0.667. The ablation likewise shows that stronger or differently timed correction is not uniformly better: the complete variants differ by only 0.067 in Vina and 3.5 percentage points in Q+V. Docking gains remain sensitive to size, compactness, and shape, and the schedule, energy terms, projection rule, and guidance strength require larger, component-matched ablations. Higher-fidelity binding assessment, protein flexibility, and stronger control of strain, synthesis, and diversity also remain open.

Interpretation also depends on the experimental unit: 10,000 molecules are nested within 100 pockets, not 10,000 independent target-level replicates. The aggregate table does not show consistency across pockets or seeds. Stronger evaluation should report paired per-pocket differences, uncertainty, seed variability, attempted-sample failure rates, and raw generated structures separately from docking-optimized structures. These analyses would separate a systematic TDGM effect from pocket composition, molecular size, and docking-pipeline sensitivity.

5 Conclusion

TDGM augments a pretrained pocket-conditioned sampler with temporal gating, pocket-direction filtering, and multi-scale physical priors. Across 100 CrossDock pockets, it changes Vina from -7.574 to -7.628, QED from 0.524 to 0.673, and mean molecular weight from 324.38 to 285.34 relative to PMDM. The diagnostic ablation and selected trajectories motivate, but do not yet isolate, the scheduling and projection mechanisms; broader controlled evaluation and experimental binding validation remain necessary.

AI USE STATEMENT

Parts of the manuscript preparation, including language refinement and editing, were assisted by generative AI tools. All scientific content, experimental design, implementation, and conclusions were developed and verified solely by the authors.

REFERENCES

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=pzpWBbnwiJ.

Lu Chen, Yan Li, Yanjie Ma, Lin Gao, and Liang Yu. Multiscale graph equivariant diffusion model for 3d molecule design, Apr 2025a. URL http://dx.doi.org/10.1126/sciadv.adv0778.

Ziqi Chen, Bo Peng, Tianhua Zhai, Daniel Adu-Ampratwum, and Xia Ning. Generating 3d small binding molecules using shape-conditioned diffusion models with guidance. *Nature Machine Intelligence*, 7:758–770, 2025b. doi: 10.1038/s42256-025-01030-w.

Julian Cremer, Tuan Le, Frank Noé, Djork-Arné Clevert, and Kristof T. Schütt. Pilot: Equivariant diffusion for pocket-conditioned de novo ligand generation with multi-objective guidance via importance sampling. *Chemical Science*, 15:14954–14967, 2024. doi: 10.1039/D4SC03523B.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021.

Zhiguang Fan, Yuedong Yang, Mingyuan Xu, and Hongming Chen. Ec-conf: A ultra-fast diffusion model for molecular conformation generation with equivariant consistency, Sep 2024. URL http://dx.doi.org/10.1186/s13321-024-00893-2.

Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9): 4200–4215, 2020. doi: 10.1021/acs.jcim.0c00411.

J. Gasteiger, J. Gross, and S. Gunnemann. Directional message passing for molecular graphs. *The Journal of Chemical Physics*, 156(20):204101, 2022. doi: 10.1063/5.0092039. URL https://doi.org/10.1063/5.0092039.

Yingqing Guo, Hui Yuan, Yukang Yang, Minshuo Chen, and Mengdi Wang. Gradient guidance for diffusion models: An optimization perspective. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Xu Han, Caihua Shan, Yifei Shen, Can Xu, Han Yang, Xiang Li, and Dongsheng Li. Training-free multi-objective diffusion model for 3d molecule generation. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=XTEkz4S9A4.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022.

Emiel Hoogeboom, Victor Garcia Satorras, Clement Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Qiaoyu Hu, Changzhi Sun, Huan He, Jiazheng Xu, Danlin Liu, Wenqing Zhang, Sumeng Shi, Kai Zhang, and Honglin Li. Target-aware 3d molecular generation based on guided equivariant diffusion. *Nature Communications*, 16(1):7928, December 2025. doi: 10.1038/s41467-025-63245-0. URL https://doi.org/10.1038/s41467-025-63245-0.

Lei Huang, Tingyang Xu, Yang Yu, Peilin Zhao, Xingjian Chen, Jing Han, Zhi Xie, Hailong Li, Wenge Zhong, Ka-Chun Wong, and Hengtong Zhang. A dual diffusion model enables 3d molecule generation and lead optimization based on target pockets. *Nature Communications*, 15 (1):2657, 2024. doi: 10.1038/s41467-024-46569-1. URL https://doi.org/10.1038/s41467-024-46569-1.

Yue Jian, Curtis Wu, Danny Reidenbach, and Aditi S. Krishnapriyan. General binding affinity guidance for diffusion models in structure-based drug design. *Journal of Chemical Information and Modeling*, 2026a. doi: 10.1021/acs.jcim.5c01166. URL https://pubs.acs.org/doi/10.1021/acs.jcim.5c01166.

Yue Jian, Curtis Wu, Danny Reidenbach, and Aditi S. Krishnapriyan. General binding affinity guidance for diffusion models in structure-based drug design. *Journal of Chemical Information and Modeling*, 66(3):1342–1352, 2026b. doi: 10.1021/acs.jcim.5c01166.

Hojung Jung, Youngrok Park, Laura Schmid, Jaehyeong Jo, Dongkyu Lee, Bongsang Kim, Se-Young Yun, and Jinwoo Shin. Conditional synthesis of 3d molecules with time correction sampler, 2024. URL https://arxiv.org/abs/2411.00551.

Cong Liu, Sharvaree Vadgama, David Ruhe, Erik Bekkers, and Patrick Forre. Clifford group equivariant diffusion models for 3d molecular generation, 2025a. URL https://arxiv.org/abs/2504.15773.

Yikai Liu, Zongxin Yu, Richard J. Lindsay, Guang Lin, Ming Chen, Abhilash Sahoo, and Sonya M. Hanson. Exendiff: An experiment-guided diffusion model for protein conformational ensemble generation. *PRX Life*, 3:023013, 2025b. doi: 10.1103/PRXLife.3.023013.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3d generative model for structure-based drug design. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6229–6239, 2021.

Marie Oestreich, Erinc Merdivan, Michael Lee, Joachim L. Schultze, Marie Piraud, and Matthias Becker. Drugdiff: small molecule diffusion model with flexible guidance towards molecular properties. *Journal of Cheminformatics*, 17(1):23, 2025. doi: 10.1186/s13321-025-01048-7. URL https://doi.org/10.1186/s13321-025-01048-7.

Yifei Qin, Xuexin Wei, Mingyuan Xu, Jiaqiang Wu, Miru Tang, Ting Ran, and Hongming Chen. Comprehensive benchmark study of diffusion-based 3d molecular generation models, Sep 2025a. URL http://dx.doi.org/10.1021/acsomega.5c05077.

Yifei Qin, Xuexin Wei, Mingyuan Xu, Jiaqiang Wu, Miru Tang, Ting Ran, and Hongming Chen. Comprehensive benchmark study of diffusion-based 3d molecular generation models. *ACS Omega*, 10(37):42760–42775, 2025b. doi: 10.1021/acsomega.5c05077.

Keyue Qiu, Yuxuan Song, Zhehuan Fan, Peidong Liu, Zhe Zhang, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. Piloting structure-based drug design via modality-specific optimal schedule. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 50619–50644. PMLR, 2025. URL https://proceedings.mlr.press/v267/qiu25g.html.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1:140022, 2014. doi: 10.1038/sdata.2014.22.

A. Sahoo and S. et al. Hanson. Exendiff: An experiment-guided diffusion model for protein conformational ensemble generation. *PRX Life*, 3:023013, 2025. doi: 10.1103/PRXLife.3.023013. URL https://journals.aps.org/prxlife/abstract/10.1103/PRXLife.3.023013.

Masami Sako, Nobuaki Yasuo, and Masakazu Sekijima. Interaction-constrained 3d molecular generation using a diffusion model enables structure-based pharmacophore modeling for drug design. *npj Drug Discovery*, 3(8), 2026. doi: 10.1038/s44386-026-00040-x.

Arne Schneuing, Yuanqi Du, Charles Harris, Arian R. Jamasb, Ilia Igashov, Weitao Du, Tom Blundell, Pietro Liò, Carla Gomes, Max Welling, Michael M. Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4:899–909, 2024. doi: 10.1038/s43588-024-00737-x. URL https://doi.org/10.1038/s43588-024-00737-x.

K. T. Schutt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Muller. SchNet –a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24): 241722, 2018. doi: 10.1063/1.5029773. URL https://doi.org/10.1063/1.5029773.

Cheng Zeng, Jirui Jin, Connor Ambrose, George Karypis, Mark Transtrum, Ellad B. Tadmor, Richard G. Hennig, Adrian Roitberg, Stefano Martiniani, and Mingjie Liu. Propmolflow: property-guided molecule generation with geometry-complete flow matching, Jan 2026. URL http://dx.doi.org/10.1038/s43588-025-00946-y.

Hao Zhang, Yang Liu, Xiaoyan Liu, Cheng Wang, and Maozu Guo. Equivariant score-based generative diffusion framework for 3d molecules, May 2024. URL http://dx.doi.org/10.1186/s12859-024-05810-w.

Peining Zhang, Daniel Baker, Minghu Song, and Jinbo Bi. Unraveling the potential of diffusion models in small-molecule generation, Jul 2025. URL http://dx.doi.org/10.1016/j.drudis.2025.104413.

SUPPLEMENTARY MATERIAL

A MATCHED-STATE EVIDENCE FOR TEMPORAL DEPENDENCE

The temporal analysis uses 120 primary states obtained from 20 pockets, with three random seeds and two independent sampling attempts per seed. Intermediate states were recorded at 10%, 30%, 50%, 70%, and 90% of generation progress. From each saved state, we evaluated four branches: an uninterrupted baseline branch, a restored numerical-shadow branch, four steps of raw collision correction, and four correction steps with a fixed norm per active step.With the network outputs held fixed, the modified and unmodified samplers agreed to within 1.19×10^{-6} Å. Across all runs, the maximum deviation of the numerical-shadow branch was 2.72×10^{-4} Å.

After ten additional updates, a raw four-step intervention applied at 10% progress retained approximately 50% of its initial clash-energy effect, compared with 93% when applied at 70% progress. The estimated paired difference in the ten-update clash-energy effect (70% minus 10%) was -2.81×10^{-5} (95% $[-4.69, -1.20] \times 10^{-5}$), where negative values indicate stronger retained clash reduction for the later intervention. Relative to the numerical-shadow branch, the final clash-atom-fraction effects were -0.0025 at 10% progress (95% CI $[-0.0058, 0.0004]$), -0.0046 at 70% (95% CI $[-0.0075, -0.0021]$), and -0.0075 at 90% (95% CI $[0.0046, 0.0102]$). At 90% progress, however, raw correction increased the internal pair-distance RMSD by 0.0073 Å relative to the numerical-shadow branch (95% CI $[0.0046, 0.0102]$).

These results provide evidence that correction persistence depends on the intervention stage. Early corrections are attenuated more strongly by subsequent denoising, whereas later corrections have larger retained effects on the final coordinates.They do not imply that guidance strength should increase monotonically, because baseline collision severity, the probability of a nonzero correction, and the resulting structural perturbation also vary across stages.

B MATCHED-BUDGET TEMPORAL-WINDOW EFFICIENCY AND ACROSS-POCKET CONSISTENCY

The candidate 50–80% window was selected using 20 development pockets and then fixed before evaluation on 50 held-out confirmation pockets. The primary comparator distributed the same nominal correction budget evenly across the full trajectory.Each of the five confirmation arms comprised 150 paired observations from 50 pockets and three fixed seeds, and all 750 trajectories completed successfully.

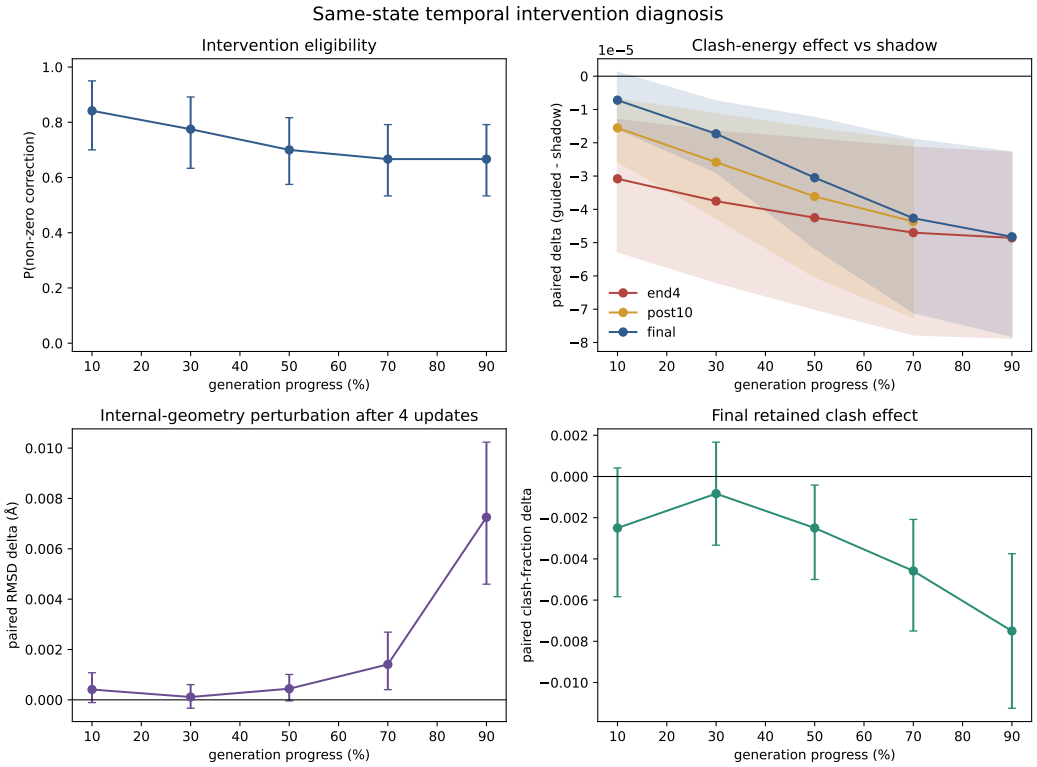

Figure 4: Temporal persistence under matched-state interventions. Error bars and shaded bands denote 95% pocket-cluster bootstrap confidence intervals.The probability of applying a nonzero correction decreases with generation progress, whereas later interventions are retained more strongly and induce larger perturbations to internal geometry.

Table 5: Matched-budget efficiency ratios for windowed versus uniform guidance. Under the efficiency definitions used here, ratios greater than 1 favor the 50–80% window.

Metric	**20-pocket development**	**50-pocket confirmation**
Per correction path	1.32 $[1.24, 1.42]$	1.39 $[1.29, 1.50]$
Per active update	4.60 $[4.27, 5.00]$	4.88 $[4.52, 5.29]$
Mean active updates	20.03 vs. 71.09	17.99 vs. 67.36

In the confirmation set, final clash energy was 4.97×10^{-5} lower under the window than under matched-budget uniform guidance. The 95% CI for the window-minus-uniform difference ranged from -8.06 $\times 10^{-5}$ to -2.60 $\times 10^{-5}$. In contrast, the clash-atom-fraction difference was +0.0020, with a 95% CI ranging from -0.0040 to 0.0080. The realized correction-path length decreased by 0.0274, with a 95% CI ranging from -0.0435 to -0.0120. The number of active updates decreased by 49.37, with a 95% CI ranging from -54.81 to -43.59. The internal pair-distance RMSD relative to the prior state decreased by 0.0508 Å, with a 95% CI ranging from -0.0730 to -0.0317. The estimated differences were 0.000 for connectivity rate (95% CI $[-0.040, 0.040]$) and +0.0067 for reconstruction rate (95% CI $[-0.0333, 0.0467]$); both confidence intervals included zero. Thus the results indicate greater intervention efficiency and lower residual clash energy. They do not, however, establish overall endpoint superiority.

B.1 Across-Pocket Consistency of Trajectory Outcomes

To assess whether the temporal result was driven by a small number of favorable cases,Figure 6 shows results for all 50 confirmation pockets. Each point represents the mean across three paired

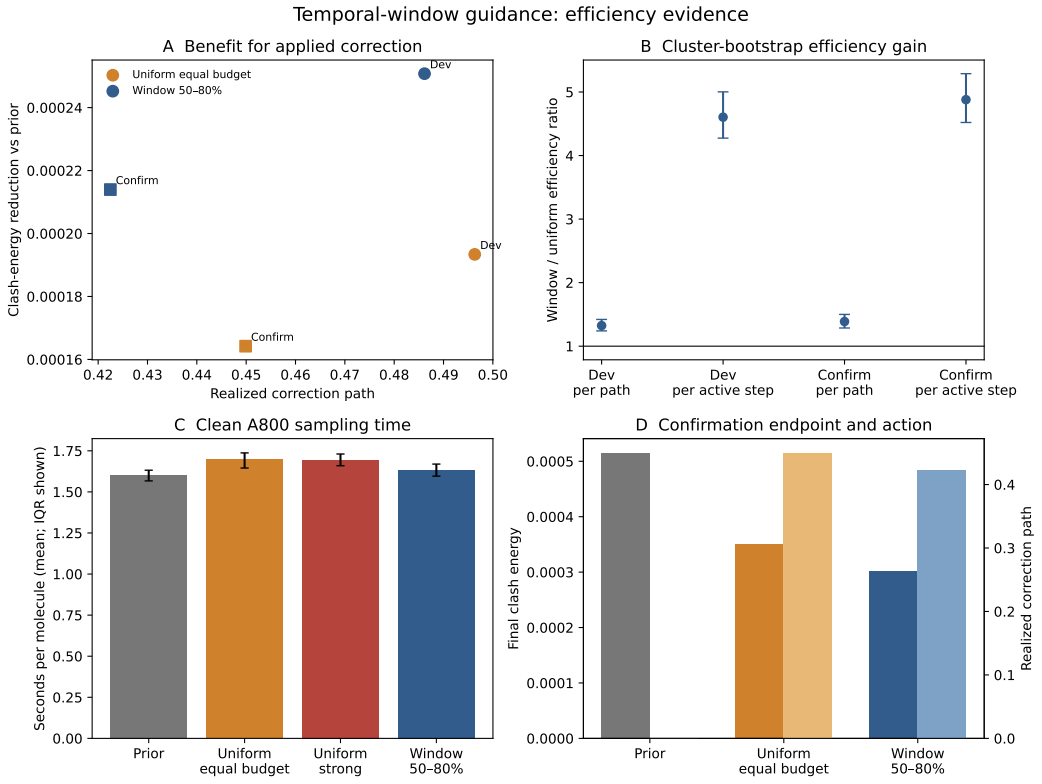

Figure 5: Efficiency of temporal-window guidance. (A) Clash-energy reduction relative to the prior state as a function of realized correction-path length.(B) Window-to-uniform efficiency ratios with 95% pocket-cluster bootstrap confidence intervals. (C) Wall-clock sampling time per molecule on an NVIDIA A800. (D) Final clash energy and realized correction-path length in the confirmation set.

seeds within one pocket, thereby assigning equal weight to pockets rather than to individual molecules. The plotted means and intervals correspond to the frozen pocket-bootstrap estimates reported above; the figure changes the presentation but not the analysis set or inferential procedure.

The distribution is not uniformly favorable: nine pockets have higher clash energy under the window and five tie; 17 pockets require a longer realized correction path; and seven show larger internal displacement. Retaining these negative cases prevents the population result from being interpreted as universal behavior. Together with the matched-state experiment in Section A, the 50-pocket distribution provides the population-level temporal evidence; the structural frames in Section F remain an illustrative visualization of one trajectory.

B.2 Clean Wall-Clock Profiling

To determine whether the reduction in active corrections translated into a wall-clock benefit, we measured per-molecule sampling time on an NVIDIA A800. Each arm was preceded by two warm-up runs; arm order was randomized, and CUDA was synchronized immediately before and after each timed run.Each arm included 60 measurements obtained from 20 pockets and three seeds.

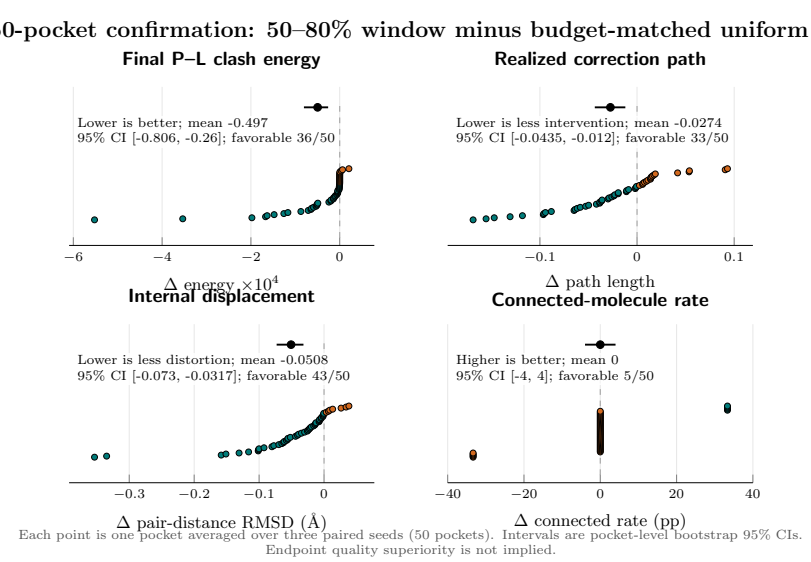

Figure 6: Pocket-level consistency of windowed guidance. Each point represents one of the 50 pocket-level effects after averaging three paired seeds. Teal and orange denote effects favorable and unfavorable to the 50–80% window, respectively; the black point and interval show the paired mean and its 95% pocket-bootstrap confidence interval. All differences are computed as windowed guidance minus budget-matched uniform guidance. The window produced lower final protein–ligand clash energy in 36 of 50 pockets, a shorter realized correction path in 33 of 50, and lower internal pair-distance displacement in 43 of 50. The estimated mean difference in connected-molecule rate was zero: 40 pockets tied, five favored the window, and five favored uniform guidance. The distribution indicates that reductions in intervention and internal structural perturbation are not confined to a few pockets, but it does not establish superior endpoint quality.

Table 6: Wall-clock sampling time per molecule on an NVIDIA A800.

Arm	Mean (s)	Median (s)
Prior only	1.6013	1.6103
Uniform equal budget	1.6965	1.7004
Uniform strong	1.6940	1.6960
50–80% window	1.6337	1.6374

The window was 0.06284 s faster than uniform equal-budget guidance (95% CI $[-0.07115, -0.05427]$), corresponding to a 3.85% speedup. The observed wall-clock reduction was consistent but modest in this profiling experiment. Because every arm still executed the same 100 network denoising steps, the 4.88-fold efficiency ratio per active correction update should not be interpreted as a 4.88-fold end-to-end speedup.

C Controlled Diagnostic of Direction Filtering

We compared unprojected and direction-filtered physical gradients to characterize the trade-off between docking score and feasibility metrics. Twenty pockets were selected by a fixed systematic rule from the ordered 100-pocket manifest before inspecting outcomes. For each pocket, we requested three samples from four arms sharing the checkpoint, atom prior, four-term energy, uniform schedule, nominal guidance scale, and sampling budget: atom-prior-only sampling, full unprojected guidance, nearest-pocket-axis projection, and the production mixed-stage pocket direction. Arm order was randomized within pocket with a fixed seed. The pocket was the resampling unit for 10,000 paired bootstrap replicates. The four arms retained 53/60, 51/60, 50/60, and 50/60 reconstructed

molecules, respectively; no shortfall was replenished, and all 204 retained molecules completed docking.

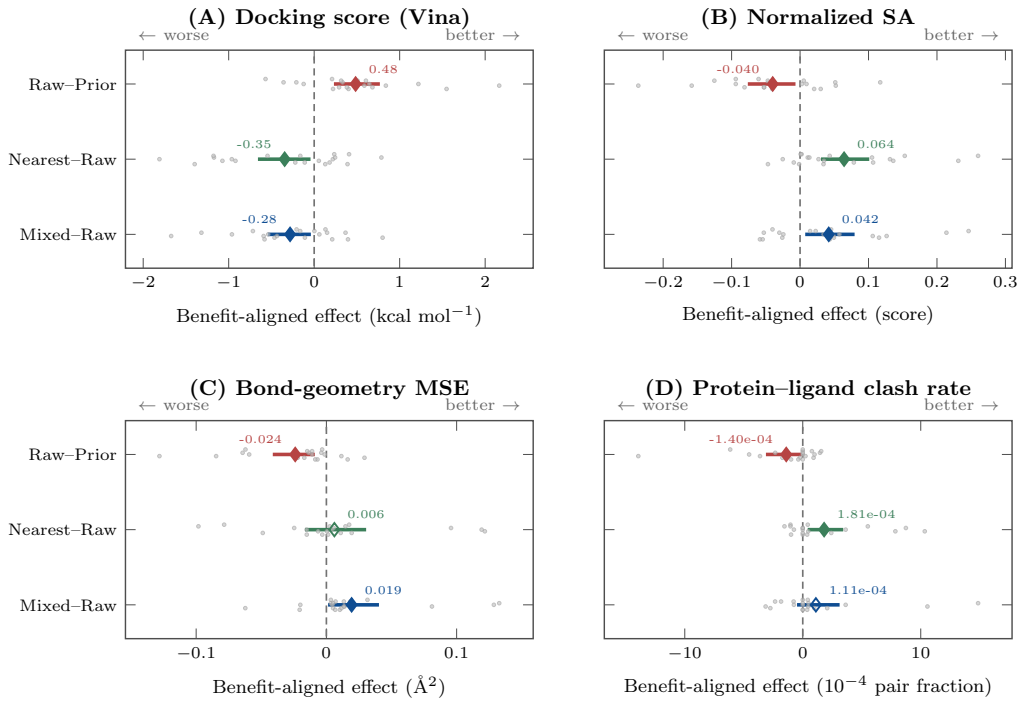

Figure 7: Trade-off under direction filtering. Effects are oriented so that positive values indicate improvement. Gray points are the 20 pocket-level effects; diamonds and horizontal lines are paired means and 95% pocket-bootstrap intervals. Filled diamonds denote intervals excluding zero. Raw guidance improves Vina but degrades normalized SA, bond geometry, and protein–ligand clash rate relative to the atom prior. Direction filtering reverses selected feasibility losses while sacrificing part of the raw Vina advantage.

Unprojected guidance improved Vina relative to atom-prior-only sampling by -0.485 (95% CI $[-0.768, -0.231]$), but reduced normalized SA by 0.040 ($[-0.076, -0.007]$), increased Bond MSE by 0.0239 ($[0.0089, 0.0412]$), and increased the PL pair-clash rate by 1.40×10^{-4} ($[0.10, 3.12] \times 10^{-4}$). Thus a more favorable docking score did not imply a uniformly more feasible raw structure, providing a direct diagnostic of the multi-objective conflict associated with unconstrained energy descent.

Direction filtering therefore mitigated different parts of the raw-gradient conflict, but did not dominate unprojected guidance on docking score. Nearest-axis projection reduced PL clashes, whereas mixed-stage projection reduced bond error; both improved normalized SA. Their Vina values were less favorable than raw guidance. Relative to atom-prior-only sampling, the average Vina differences were -0.139 for nearest projection (95% CI $[-0.468, +0.165]$) and -0.203 for mixed-stage projection ($[-0.415, +0.016]$), so preservation of a docking benefit was not conclusive. The bounded interpretation is that pocket-direction filtering regularizes selected feasibility dimensions while exposing a feasibility–docking trade-off.

This experiment does not isolate direction from realized update magnitude because projection changes the gradient norm, and it uses one checkpoint seed rather than independent-seed confirmation. An equal-norm raw control is therefore required for a direction-only causal claim. The experiment is not used to assign the 100-pocket aggregate gain to projection alone.

D FROZEN STRENGTH CONFIRMATION AND PENETRATION PROFILE

Window strengths 0.020, 0.025, 0.030, and 0.035 were screened on 20 development pockets. The prespecified rule required realized correction path within $\pm 10\%$ of the strong uniform arm and no more than a two-percentage-point decrease in connectivity or reconstruction, followed by selection of the lowest clash energy. Strength 0.035 passed and was frozen before evaluation on 50 confirmation pockets.

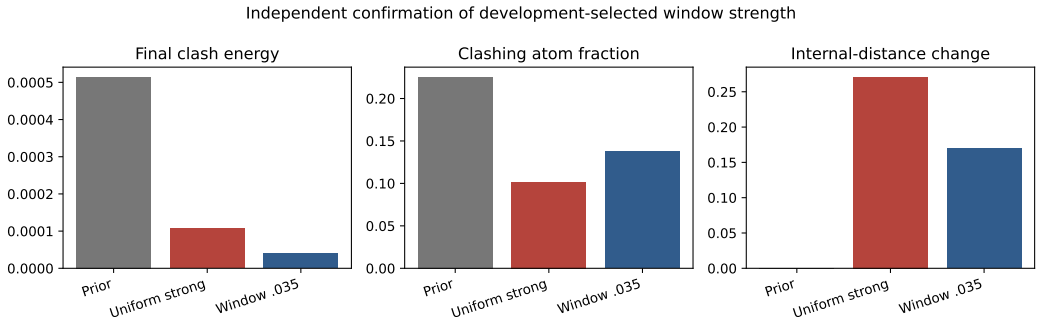

Figure 8: Confirmation of window strength. The window lowers clash energy and internal-distance change but has a higher fraction of atoms below the single 1.8 Å cutoff than uniform strong guidance.

Table 7: Frozen 50-pocket comparison of the 0.035 window against strong uniform guidance. Paired differences are window minus uniform.

Metric	Uniform strong	Window 0.035	Paired difference [95% CI]
Clash energy	1.074×10^{-4}	4.001×10^{-5}	-6.74×10^{-5} $[-1.09 \times 10^{-4}, -3.32 \times 10^{-5}]$
Atom fraction < 1.8 Å	0.1020	0.1377	$+0.0357$ $[0.0250, 0.0470]$
Connected rate	0.6000	0.6067	$+0.0067$ $[-0.0600, 0.0733]$
Reconstruction rate	0.6200	0.6267	$+0.0067$ $[-0.0600, 0.0733]$
Correction path	1.3545	1.1892	-0.1653 $[-0.2162, -0.1159]$
Internal pair RMSD (Å)	0.2713	0.1698	-0.1016 $[-0.1352, -0.0709]$

The apparent disagreement between lower clash energy and a higher < 1.8 Å fraction is resolved by scanning the distance threshold rather than relying on one cutoff.

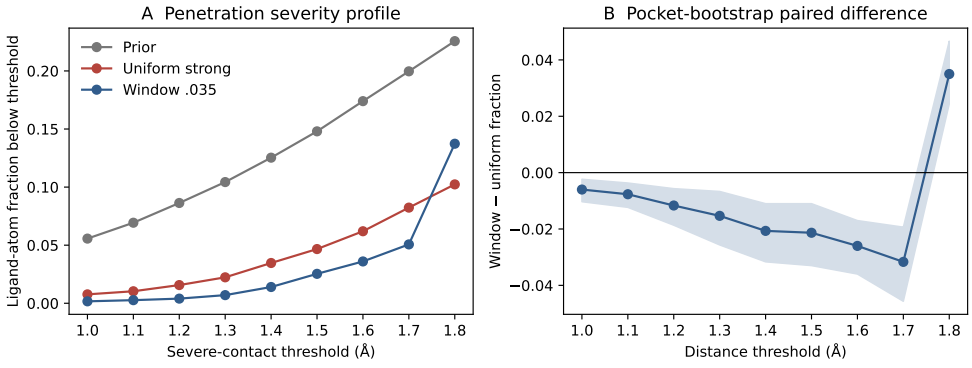

Figure 9: Penetration across distance thresholds. The 0.035 window reduces the fraction of atoms below every threshold from 1.0 to 1.7 Å, but reverses at the 1.8 Å hard cutoff. Bands in the paired-difference panel are 95% pocket-cluster bootstrap intervals.

Relative to uniform strong guidance, the window changes penetration fractions by -1.17, -2.07, -2.60, and -3.17 percentage points at 1.2, 1.4, 1.6, and 1.7 Å, respectively, but by $+3.50$ percentage

points at 1.8 Å. The result is therefore not a uniform reduction of all contacts: severe penetration decreases, while shallow contacts accumulate just inside the implemented cutoff.

E Geometry and Energy Diagnostics on the 100-Pocket Benchmark

The separate 100-pocket benchmark evaluates computational feasibility diagnostics that are not captured by docking score, QED, or Lipinski counts. The retained PMDM and TDGM samples comprise 985 and 964 molecules, respectively. Because these are not paired mechanism interventions, the table describes system-level differences and cannot isolate scheduling or projection.

Table 8: Geometry and energy diagnostics on 100 CrossDock pockets. Lower is better for Bond MSE, clash rates, MMFF median, and the high-energy tail.

Method	n	**Bond MSE**	**Lig. clash**	**PL clash**	**Min PL dist.**	**MMFF med.**	**MMFF** > 100
PMDM	985	0.0861	0.0160	5.34×10^{-4}	1.757	24.58	29.5%
TDGM	964	0.0711	0.0150	4.63×10^{-4}	1.792	25.43	25.5%

TDGM has lower Bond MSE and clash rates, a larger minimum protein–ligand distance, and a smaller fraction of MMFF energies above 100. Its median MMFF energy is slightly higher (25.43 versus 24.58), so the force-field diagnostic is mixed rather than uniformly improved. These computational diagnostics do not establish experimental stability or binding affinity.

Because retained molecule counts vary from 7–10 per pocket for PMDM and 6–10 per pocket for TDGM, we also aggregate each metric within pocket and bootstrap the 100 paired pocket differences. This prevents pockets with more retained molecules from receiving greater inferential weight.

Table 9: Paired 100-pocket diagnostics. Differences are TDGM minus PMDM; intervals use 10,000 paired pocket bootstrap resamples.

Pocket-aggregated metric	**PMDM**	**TDGM**	**Paired difference [95% CI]**
Bond MSE	0.0870	0.0724	$-0.0146\ [-0.0207, -0.0086]$
Ligand clash rate	0.0161	0.0151	$-0.00091\ [-0.00249, 0.00071]$
Protein–ligand clash rate	5.46×10^{-4}	4.82×10^{-4}	$-6.44 \times 10^{-5}\ [-1.14 \times 10^{-4}, -1.72 \times 10^{-5}]$
Minimum protein–ligand distance (Å)	1.750	1.779	$+0.029\ [0.007, 0.051]$
Fraction with MMFF strain > 100	0.299	0.261	$-0.038\ [-0.062, -0.015]$
Retained fraction of nominal ten	0.985	0.964	$-0.021\ [-0.035, -0.008]$

The last row is an availability diagnostic, not a generation-validity estimate, because the archived pipeline does not preserve a complete attempt-level failure ledger. The lower retained fraction for TDGM is therefore reported rather than silently omitted. The paired analysis supports lower Bond MSE, protein–ligand clash rate, and high-strain tail together with a larger minimum protein–ligand distance on this retained set; the ligand-only clash interval crosses zero. These results remain descriptive package-level comparisons.

F Illustrative Structural Trajectory

The main text reports five frames for TDGM-096-3. The denser 11-frame visualization below documents the same selected trajectory at 10% progress intervals. It is placed after the multi-pocket mechanism analyses so that a visually salient case is not mistaken for the primary evidence.

The frames show a continuous change for this example, but a selected case cannot establish a general temporal mechanism. Only this case has archived stepwise coordinates suitable for a consistent protein–ligand structural rendering. The matched-state intervention and complete 50-pocket consistency analysis in Sections A–B provide the population-level evidence; a future structural gallery requires prospectively saved stepwise coordinates for every displayed pocket and must include failures rather than selecting favorable trajectories.

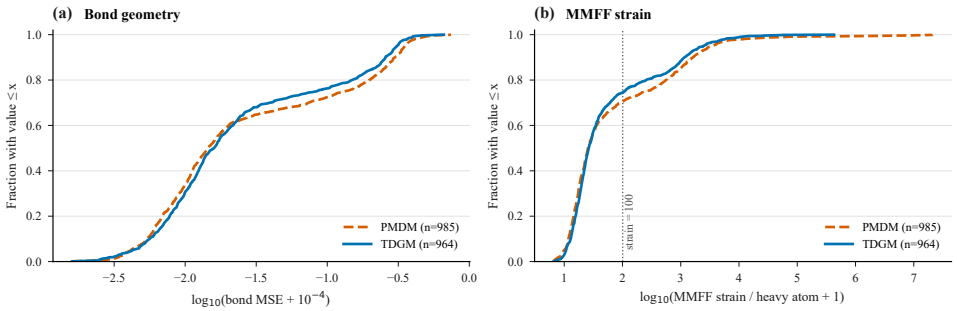

Figure 10: Geometry and strain diagnostics. Distribution tails complement the aggregate diagnostics in Table 8.

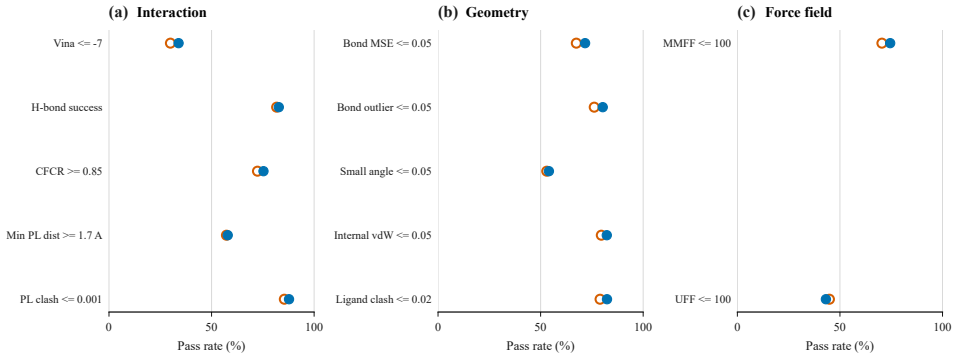

Figure 11: Metric-family pass rates and endpoint shifts. The display is descriptive and does not isolate individual TDGM modules.

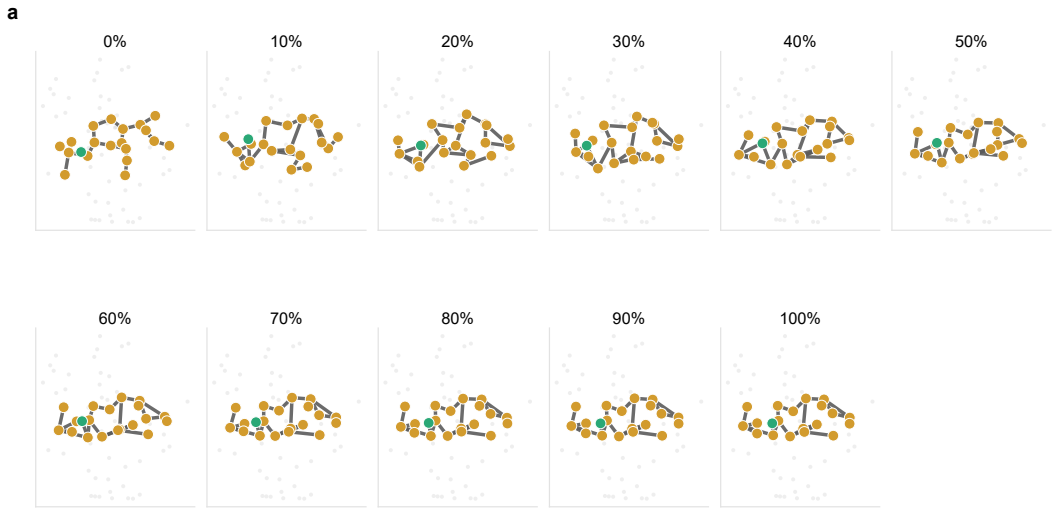

Figure 12: Illustrative denoising trajectory at 10% progress intervals. RMSD and bond-error changes are measured relative to this trajectory's final generated structure, not to an experimental ligand pose.