Harmonic Torsional Diffusion for Protein-Ligand Flexible Docking

Maksim Zhdanov [1 2] **Pavel Strashnov** [2] **Vladislav Kurenkov** [1 2]

Abstract

Molecular docking requires reasoning jointly about ligand pose and protein flexibility. Most diffusion-based docking models predict torsional updates with generic Euclidean heads that ignore the periodic geometry of angular variables. This mismatch is especially limiting in flexible docking, where ligand conformations and pocket side chains co-adapt to form the bound complex. Here, we introduce Harmony, a harmonic torsional diffusion framework for flexible protein-ligand docking. Harmony parameterizes ligand and side-chain torsional score fields as derivatives of learned harmonic potentials on the circle, whose noise-level dependence is supplied analytically by the heat semigroup of variance-exploding diffusion on the torus. This construction makes periodicity explicit and gives the model a frequency-aware inductive bias over rotameric motion. On the PDBBind benchmark, Harmony improves ligand pose accuracy and pocket all-atom reconstruction over recent flexible docking methods. On PoseBusters, it improves the physical validity of generated complexes. Case studies on EBNA1 and KRAS G12D illustrate the method's behavior on a polar and a shallow binding site, respectively. Together, these results indicate that aligning the score parameterization with the geometry of the diffusion process is a simple and effective lever for improving flexible docking.

1. Introduction

Understanding binding phenomena between small molecules and proteins is central to modern drug discovery and computational methods aim to characterize these interactions through molecular docking (Pagadala et al., 2017). The field has evolved substantially, beginning with classical search-based strategies and more recently incorporating deep learning approaches. Despite these advances, a large portion of existing methods operates under the simplifying assumption that protein structures remain fixed during binding.

This assumption contrasts with the inherently dynamic nature of proteins, which frequently adopt different conformations upon ligand association (Weikl & Paul, 2014). To address this, computational strategies that account for structural variability are typically divided into co-folding and flexible docking approaches (Abramson et al., 2024; Corso et al., 2025). Co-folding attempts to determine the bound configuration of both protein and ligand simultaneously, treating the complex as a single prediction problem. In contrast, flexible docking focuses on capturing the conformational adjustments that occur between pre-existing unbound and bound states, offering advantages in efficiency, interpretability and controllability.

Even so, achieving reliable accuracy with flexible docking remains a challenge. Traditional search-based methods are hindered by the expanded dimensionality introduced by protein flexibility, making exhaustive exploration impractical (McNutt et al., 2021; Koes et al., 2013). Deep learning-based methods attempt to mitigate this by integrating diffusion or flow-based frameworks with or without additional flexibility (Corso et al., 2022; Plainer et al., 2023; Corso et al., 2025). However, they rely on the basic Riemannian generative machinery (Jing et al., 2022; De Bortoli et al., 2022; Chen & Lipman, 2023), which does not utilize the rich geometric structure inherent in conformational degrees-of-freedom. In particular, protein side chains demonstrate specific conformations that are called rotamers, arising from rotations around χ dihedral (torsional) angles (Branden & Tooze, 2012). The exact configuration of side chains is largely explained by steric repulsion between atoms whose spatial arrangement forms torsional angles. Because of this, most amino acids prefer to sit in discrete states in which they are positioned by some optimal dihedral angle value relative to the next groups. Importantly, ligand binding can further shift side chain rotamers by means of induced fit, reflecting a coupled relationship in which both ligand conformation and protein side-chain orientations adapt to optimize interactions (Gaudreault et al., 2012). Deep learning models for flexible docking should be able to naturally learn those multi-frequency motion features (Bahar et al.,

[1]dunnolab.ai [2]AXXX. Correspondence to: Maksim Zhdanov <dwdcodes@gmail.com>.

Accepted at the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026)

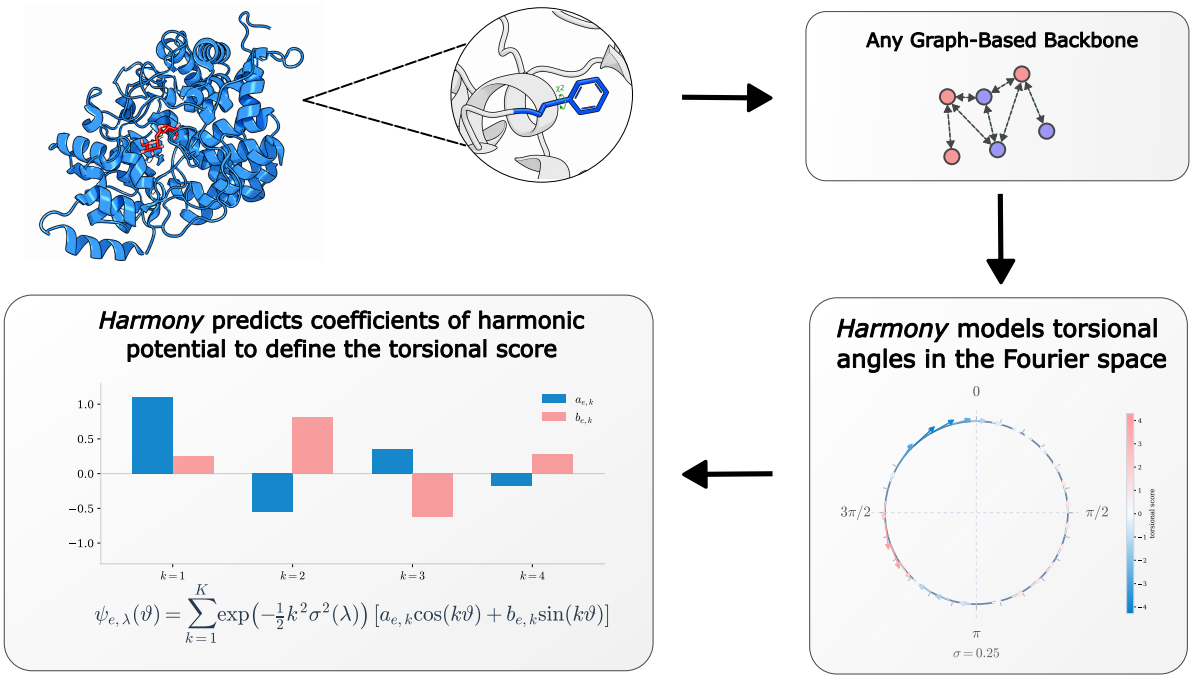


Figure 1. Overview of *Harmony*. Starting from a representation produced by any graph-based neural backbone, *Harmony* replaces direct torsion regression with a structured periodic parameterization. Instead of predicting a single angular update, the model learns harmonic coefficients that define a Fourier potential, whose derivative yields the torsional score field for ligand and flexible side-chain dihedrals. This construction makes periodicity explicit and enables the model to naturally represent difficult torsional configurations. See more in Section 4.3.

2010) to enable higher structural validity of predicted poses.

To address those challenges we propose *Harmonic Torsional Diffusion* (in experimental section coined *Harmony*), where we parametrize torsional angles of ligands and protein side chains in a Fourier basis, which comes with a clean analytical score function formulation and efficient noising structure which gradually suppresses frequency motion components coming from high to low harmonic variations.

We study this parametrization from the formal and empirical perspectives. We demonstrate that Harmony improves over recent generative methods for flexible docking on PDBBind benchmark (Liu et al., 2017) by ∼24.5% points while following their training setup and model architecture. Moreover, on PoseBusters set (Buttenschoen et al., 2024), Harmony shows improvements in chemical validity of generated poses over the large portion of baselines. To demonstrate specific mechanism of our method empirically we conduct two case studies: EBNA1 and KRAS G12D. Additionally, we provide ablation studies on design choices behind *Harmony*.

In summary, we propose Harmonic Torsional Diffusion, a new generative framework for flexible docking. We prove that it naturally represents the specific degrees-of-freedom central to molecular binding.

The source code is publicly available.[1]

2. Related Work

Flexible Molecular Docking. Flexible docking is typically formulated under the assumption that the unbound conformation of a protein (apo state) is available, with the objective of predicting the structural adjustments that lead to the bound complex (holo state). Classical docking approaches tackle this problem by defining an energy or scoring function and exploring the space of possible ligand poses to identify configurations that minimize this objective (Friesner et al., 2004; Thomsen & Christensen, 2006; Trott & Olson, 2010). In principle, protein flexibility, crucial to real-world applications (Teague, 2003), can be incorporated by extending the search space to include additional degrees of freedom, such as side-chain torsional angles within the binding pocket. In practice, however, this significantly enlarges the search space, making it difficult to efficiently identify optimal joint configurations and often resulting in physically inconsistent or suboptimal poses.

More recently, deep learning-based approaches have been introduced to address these limitations. Methods such as

[1] https://github.com/dunnolab/harmony

DiffDock-Pocket (Plainer et al., 2023), Re-Dock (Huang et al., 2024), FlexDock (Corso et al., 2025) extend generative frameworks, ranging from standard diffusion models to diffusion bridges and flow-based formulations, to account for flexibility in both ligands and protein binding pockets. Within this line of work, some methods restrict flexibility to side chains while treating the protein backbone as fixed, whereas others attempt to capture full structural variability, including backbone motion. Both perspectives can be justified depending on the biological context. Recently, SigmaDock (Prat et al., 2025) also explored the idea that the torsional module for docking is not optimal by reframing generative process with molecular fragmentation. However, they only focus on rigid pocket-based docking which is an easier subtask.

The approach presented here departs from these existing methods by introducing a new generative framework specifically designed to naturally model torsional motion in both ligands and proteins. This *harmonic torsional diffusion* formulation offers advantages in both efficiency and predictive performance. It is applicable in both pocket-based and blind docking settings and can accommodate either rigid or flexible protein representations. In the present work, the focus is restricted to side-chain flexibility, with a more detailed description of the task formulation provided in Section 3.

Diffusion Models. Score-based diffusion models (Song et al., 2020) define a continuous diffusion process $dx = f(x,t)dt + g(t)dw$ that transports the data distribution p_0 in a simple prior p_T. This has the corresponding reverse SDE $dx = [f(x,t) - g(t)^2 \nabla_x \log p_t(x)]dt + g(t)dw$ where only the score $\nabla_x \log p_t(x)$ is unknown. Using denoising score matching (Vincent, 2011), one can learn $s(x,t) \approx \nabla_x \log p_t(x)$ and use it to run the reverse SDE to obtain samples from the data distribution. We base *Harmony's* noising process on Variance-Exploding (VE) SDE and further demonstrate its relation to the circular nature of torsional angles.

3. Preliminaries

All-Atom Ligand–Protein Representation. We represent each complex as a heterogeneous graph

$$\mathcal{G} = \{\mathcal{V} \coloneqq (\mathcal{V}^l, \mathcal{V}^b, \mathcal{V}^a),\ \mathcal{E} \coloneqq (\mathcal{E}^l, \mathcal{E}^b, \mathcal{E}^a, \mathcal{E}^{lb}, \mathcal{E}^{la}, \mathcal{E}^{ab})\},$$

where $\mathcal{V}^l$ denotes ligand atoms, $\mathcal{V}^b$ residue-level receptor backbone nodes at C_α positions, and $\mathcal{V}^a$ receptor atoms including both backbone and side-chain atoms. Ligand nodes carry standard atom-level chemical descriptors, residue nodes encode amino-acid identity together with a pretrained ESM-2 embedding (Lin et al., 2022), and receptor atom nodes encode residue identity and atom-type information. The edge sets $\mathcal{E}^l, \mathcal{E}^b, \mathcal{E}^a$ capture intra-graph ligand, residue, and atom interactions, while $\mathcal{E}^{lb}, \mathcal{E}^{la}, \mathcal{E}^{ab}$ couple ligand atoms to receptor residues, ligand atoms to receptor atoms, and receptor atoms to their parent residues. This multi-scale construction allows the model to reason jointly over ligand chemistry, residue-level geometry, and full atomic structure. Detailed feature definitions and graph construction are given in Appendix D.1.

Pocket-Based Flexible Docking. To focus modeling on the binding interface, we replace the full receptor with a pocket-centered subgraph $\mathcal{G}_{\text{pocket}}$ containing the ligand together with nearby receptor residues and atoms, following standard pocket-reduction practice (Corso et al., 2025). We keep the protein backbone fixed and allow motion only in pocket side chains, since side-chain rearrangements near the binding site account for much of the local conformational adaptation required for docking (Miao & Cao, 2016; Clark et al., 2019; Alberts et al., 2005). The docking state is therefore parameterized as

$$\mathbf{z} = (\mathbf{t}, \mathbf{R}, \boldsymbol{\tau}, \boldsymbol{\chi}) \in \mathcal{M} = \mathbb{R}^3 \times SO(3) \times \mathbb{T}^{m_l} \times \mathbb{T}^{m_s},$$

where $\mathbf{t} \in \mathbb{R}^3$ and $\mathbf{R} \in SO(3)$ denote ligand translation and rotation, $\boldsymbol{\tau} \in \mathbb{T}^{m_l}$ denotes ligand torsions, and $\boldsymbol{\chi} \in \mathbb{T}^{m_s}$ denotes flexible pocket side-chain torsions. Thus, docking is modeled on a product manifold with Euclidean, rotational, and toroidal components, and the network jointly predicts ligand pose and pocket side-chain arrangement from $\mathcal{G}_{\text{pocket}}$. Exact pocket selection details are deferred to Appendix C.2.

4. Harmonic Torsional Diffusion

In this section, we firstly build the foundation of the *Harmonic Torsional Diffusion* (*Harmony*) and then derive explicit parametrization and training for the model. All the proofs for the relevant propositions can be found in Appendix A.

4.1. Variance-Exploding SDE

We adopt the variance-exploding (VE) diffusion SDE framework as the probabilistic foundation of our method. In VE diffusion, the forward process perturbs data through pure stochastic diffusion, without any deterministic drift term. In Euclidean space, the forward SDE takes the form

$$\mathrm{d}\mathbf{z}_\lambda = g(\lambda)\,\mathrm{d}\mathbf{w}_\lambda,$$

where $\mathbf{w}_\lambda$ is standard Brownian motion and $g(\lambda)$ is a time-dependent diffusion coefficient. Unlike variance-preserving (VP) SDE, VE-SDE diffusion models do not contract the state toward the origin during the forward process. Instead, the variance increases monotonically with time, gradually transforming the data distribution into a prior distribution.

The corresponding marginal perturbation kernel is Gaussian, $p_\lambda(\mathbf{z}_\lambda \mid \mathbf{z}_0) = \mathcal{N}\big(\mathbf{z}_0, \sigma(\lambda)^2\mathbf{I}\big)$, where $\sigma(\lambda)$ is the noise

scale induced by the diffusion coefficient $g(\lambda)$. In practice, we use an exponential noise schedule, $\sigma(\lambda) = \sigma_{\min}^{1-\lambda}\sigma_{\max}^{\lambda}$.

A key property of the VE forward process is that its density evolution satisfies a heat equation as demonstrated in Theorem 4.1. The proposition above suggests a natural parameterization for torsional variables. Since each torsion angle lives on the circle $\mathbb{T}$ and the VE forward process acts as heat flow, the appropriate basis is given by the eigenfunctions of the Laplace-Beltrami operator on $\mathbb{T}$ which are the Fourier modes. We elaborate on the main idea in the following subsections.

Proposition 4.1 (VE-SDE as heat flow, cf. Sarkar (2026))**.** *Let the forward process be defined by the driftless stochastic differential equation*

$$\mathrm{d}\mathbf{z}_\lambda = g(\lambda)\,\mathrm{d}\mathbf{w}_\lambda, \qquad \mathbf{z}_0 \sim p_0,$$

where $\mathbf{w}_\lambda$ is standard Brownian motion and $g(\lambda)$ is a time-dependent diffusion coefficient. Then the density $p_\lambda(\mathbf{z})$ of $\mathbf{z}_\lambda$ satisfies the Fokker–Planck equation

$$\frac{\partial p_\lambda(\mathbf{z})}{\partial \lambda} = \frac{g(\lambda)^2}{2}\,\Delta p_\lambda(\mathbf{z}),$$

that is, a heat equation with time-dependent diffusivity

$$\nu(\lambda) = \frac{g(\lambda)^2}{2}.$$

4.2. Product Space Diffusion

We model flexible docking with a variance-exploding (VE) diffusion process on the product space of ligand rigid-body motion and periodic torsional variables. Since the protein backbone is kept fixed and only pocket side chains are allowed to move, the forward process as defined according to $\mathbf{z}$, perturbs the translation, rotation, ligand torsions and pocket side-chain torsions independently at noise level $\lambda \in [0, 1]$:

$$q_\lambda(\mathbf{z}_\lambda \mid \mathbf{z}_0) = q_\lambda^{\mathrm{tr}}(\mathbf{t}_\lambda \mid \mathbf{t}_0)\, q_\lambda^{\mathrm{rot}}(\mathbf{R}_\lambda \mid \mathbf{R}_0)\, q_\lambda^{\mathrm{tor}}(\boldsymbol{\tau}_\lambda \mid \boldsymbol{\tau}_0)\, q_\lambda^{\mathrm{sc}}(\boldsymbol{\chi}_\lambda \mid \boldsymbol{\chi}_0).$$

where each component is corrupted with an exponentially increasing noise scale

$$\sigma_u(\lambda) = \sigma_{u,\min}^{1-\lambda}\sigma_{u,\max}^{\lambda}, \qquad u \in \{\mathrm{tr}, \mathrm{rot}, \mathrm{tor}, \mathrm{sc}\}.$$

For ligand translation, we apply Gaussian perturbations in Euclidean space,

$$\mathbf{t}_\lambda = \mathbf{t}_0 + \sigma_{\mathrm{tr}}(\lambda)\,\boldsymbol{\epsilon}_{\mathrm{tr}}, \qquad \boldsymbol{\epsilon}_{\mathrm{tr}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3).$$

For ligand rotation, we perturb the clean orientation on $SO(3)$ with isotropic rotational noise. In practice, this perturbation is represented in axis-angle form and applied through the exponential map,

$$\mathbf{R}_\lambda = \mathrm{Exp}(\boldsymbol{\omega}_\lambda)\,\mathbf{R}_0, \qquad \boldsymbol{\omega}_\lambda \sim \mathrm{IGSO}(3; \sigma_{\mathrm{rot}}(\lambda)).$$

For ligand torsions and pocket side-chain torsions, we use wrapped Gaussian perturbations on the torus,

$$\boldsymbol{\tau}_\lambda = \mathrm{wrap}(\boldsymbol{\tau}_0 + \sigma_{\mathrm{tor}}(\lambda)\,\boldsymbol{\epsilon}_{\mathrm{tor}}), \qquad \boldsymbol{\epsilon}_{\mathrm{tor}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

$$\boldsymbol{\chi}_\lambda = \mathrm{wrap}(\boldsymbol{\chi}_0 + \sigma_{\mathrm{sc}}(\lambda)\,\boldsymbol{\epsilon}_{\mathrm{sc}}), \qquad \boldsymbol{\epsilon}_{\mathrm{sc}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

where $\mathrm{wrap}(\cdot)$ maps angles back to $(-\pi, \pi]$.

The network is trained to predict the score of each factor of the forward noising distribution. Reverse-time sampling then jointly denoises these components to recover the bound ligand pose together with the binding-pocket side-chain arrangement according to backward process (Anderson, 1982).

4.3. Harmonic Parametrization of Torsional Angles

Overall design of *Harmony* can be seen in Figure 1. The VE-SDE viewpoint above motivates a natural representation for torsional variables. Since ligand torsions $\boldsymbol{\tau}$ and side-chain torsions $\boldsymbol{\chi}$ take values on a torus, each individual torsional degree of freedom lies on the circle $\mathbb{T}$. Two facts make this geometry tractable.

Proposition 4.2 (Fourier eigenbasis of the circle, Stein & Shakarchi (2003))**.** *Let $\Delta_{\mathbb{T}}$ denote the Laplace-Beltrami operator on the circle $\mathbb{T}$. Then, for every integer $k \geq 1$, the functions $\cos(k\vartheta)$ and $\sin(k\vartheta)$ are eigenfunctions of $\Delta_{\mathbb{T}}$ with eigenvalue $-k^2$, i.e.*

$$\Delta_{\mathbb{T}} \cos(k\vartheta) = -k^2 \cos(k\vartheta),$$
$$\Delta_{\mathbb{T}} \sin(k\vartheta) = -k^2 \sin(k\vartheta).$$

Consequently, for any finite Fourier expansion

$$\psi(\vartheta) = \sum_{k=1}^{K} \big[a_k \cos(k\vartheta) + b_k \sin(k\vartheta)\big],$$

the heat semigroup generated by $\Delta_{\mathbb{T}}$ acts diagonally:

$$\exp\left(\frac{\sigma^2}{2}\Delta_{\mathbb{T}}\right)\psi(\vartheta) = \sum_{k=1}^{K} \exp\left(-\frac{1}{2}k^2\sigma^2\right) \big[a_k \cos(k\vartheta) + b_k \sin(k\vartheta)\big].$$

Proposition 4.3 (Wrapped Gaussian noising is heat flow on $\mathbb{T}$, Mardia & Jupp (2009, §3.5.7))**.** *Let $\vartheta_\lambda = \mathrm{wrap}(\vartheta_0 + \sigma(\lambda)\,\epsilon)$ with $\epsilon \sim \mathcal{N}(0, 1)$. Then the conditional density of ϑ_λ given ϑ_0 is the wrapped normal*

$$p_\lambda(\vartheta \mid \vartheta_0) = \sum_{n\in\mathbb{Z}} \frac{1}{\sqrt{2\pi\sigma^2(\lambda)}} \exp\left(-\frac{(\vartheta - \vartheta_0 - 2\pi n)^2}{2\sigma^2(\lambda)}\right),$$

which coincides with the heat kernel on $\mathbb{T}$ at time $\sigma^2(\lambda)/2$. Equivalently, $p_\lambda(\cdot \mid \vartheta_0) = \exp\left(\frac{\sigma^2(\lambda)}{2}\Delta_{\mathbb{T}}\right)\delta_{\vartheta_0}$, so the marginal density satisfies $\partial_\lambda p_\lambda = \frac{g(\lambda)^2}{2}\Delta_{\mathbb{T}}\, p_\lambda$.

Let ϑ denote one component of either $\boldsymbol{\tau}$ or $\boldsymbol{\chi}$ associated with a rotatable bond e. For each such bond, the message-passing backbone produces a local bond representation $\mathbf{u}_e$, from which *Harmony* predicts harmonic coefficients via

$$[a_{e,1}, b_{e,1}, \ldots, a_{e,K}, b_{e,K}] = \mathbf{g}_\theta(\mathbf{u}_e),$$

where K is the number of retained Fourier modes. These coefficients define a latent harmonic function on the circle,

$$\psi_e(\vartheta) = \sum_{k=1}^{K} \left[a_{e,k}\cos(k\vartheta) + b_{e,k}\sin(k\vartheta)\right].$$

By Propositions 4.1-4.3, the VE forward process on each torsional coordinate acts as heat flow on $\mathbb{T}$, and Proposition 4.2 diagonalizes this action in the Fourier basis. The smoothed harmonic function at noise level λ is therefore

$$\psi_{e,\lambda}(\vartheta) = \sum_{k=1}^{K} \exp\left(-\frac{1}{2}k^2\sigma^2(\lambda)\right) \left[a_{e,k}\cos(k\vartheta) + b_{e,k}\sin(k\vartheta)\right]. \tag{1}$$

Thus, VE diffusion induces a frequency-dependent damping, where high-frequency angular structure is suppressed at large noise, while finer modes gradually re-emerge as $\lambda \to 0$ as in Figure 2.

Harmony predicts the torsional score field analytically by differentiating the heat-smoothed harmonic expansion with respect to the angle:

$$\begin{aligned} s_e(\vartheta, \lambda) &= \partial_\vartheta \psi_{e,\lambda}(\vartheta) \\ &= \sum_{k=1}^{K} k \exp\left(-\frac{1}{2}k^2\sigma_u^2(\lambda)\right) \\ &\quad \left[-a_{e,k}\sin(k\vartheta) + b_{e,k}\cos(k\vartheta)\right]. \end{aligned}$$

Applying this construction independently to all ligand and side-chain rotatable bonds yields the torsional score components associated with $\boldsymbol{\tau}$ and $\boldsymbol{\chi}$.

This parameterization is crucial because the model does not need to learn an arbitrary score field from scratch. Instead, the dependence on the VE noise level is built in explicitly through the heat-semigroup damping factor $\exp(-\frac{1}{2}k^2\sigma_u^2(\lambda))$, which is exactly how the forward process transforms Fourier modes on $\mathbb{T}$. As a result, the representation remains faithful to the noising process across all noise levels, preserves periodicity by construction and naturally captures multimodal torsional landscapes such as ligand conformers and side-chain rotamers.

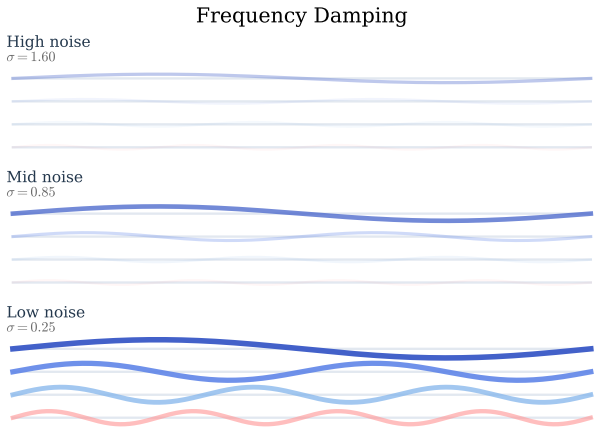


Figure 2. Frequency damping effect of VE noising. As noise level increases, high-frequency components saturate faster, but low-frequency modes remain. Thus, *Harmony* learns torsional motion naturally by filtering different scales of molecular motion during training.

4.4. Confidence and Relaxation Modules

Confidence Head. We use a lightweight confidence head attached to the shared docking encoder rather than a separate reranking model. The head predicts a scalar confidence score for each sampled pose and is supervised with protein-ligand interface lDDT (PLI-lDDT) (Lu et al., 2024), which measures local receptor–ligand contact accuracy. To emphasize near-final geometries, confidence is trained with a time-weighted regression loss that upweights low-noise samples. More details in Section D.3.

Relaxation. We do not apply post hoc relaxation in the present study. This choice allows us to isolate the quality of the manifold docking model itself and evaluate the generated complexes directly. In principle, *Harmony* can be combined with any downstream relaxation procedure, including separately trained relaxation models similar to FlexDock (Corso et al., 2025) or external tools such as GNINA (McNutt et al., 2021).

Model Architecture and Training. *Harmony* uses an e3nn (Geiger & Smidt, 2022) heterogeneous $SE(3)$-equivariant graph neural network over the ligand atom graph, the receptor C_α graph, and the receptor atom graph in the binding pocket. We train the model with the standard diffusion score-matching objective as in (Plainer et al., 2023). Training uses OpenEquivariance (Bharadwaj et al., 2025) kernels for Clebsch-Gordan tensor products, reducing runtime from roughly 4.5 days to 2 days on 8 H100 80GB GPUs (see Appendix D.3).

Table 1. Top-1 PDBBind ESMFold docking performance. *Harmony* samples 10 complexes and selects the one with the highest confidence. Best results are shown in bold. N.A. indicates missing results for baselines whose output structures were not available. Runtime for *Harmony* is reported both for the standard Clebsch–Gordan tensor product implementation and for the OpenEquivariance CUDA-kernel version used in our base model.

Method	Ligand RMSD		All-Atom	Runtime (s)
	% < 2 Å ↑	Med. Å ↓	% < 1 Å ↑	
SMINA (rigid)	6.6	7.7	N.A.	258
SMINA	3.6	7.3	5.2	1914
GNINA (rigid)	6.7	7.1	N.A.	260
GNINA	8.4	7.9	4.5	1575
DiffDock-Pocket	41.8	2.5	32.4	17
ReDock	39.0	2.5	39.8	15
FlexDock	39.7	2.5	41.7	11
Harmony	**64.2**	**1.9**	**47.3**	**5 (11)**

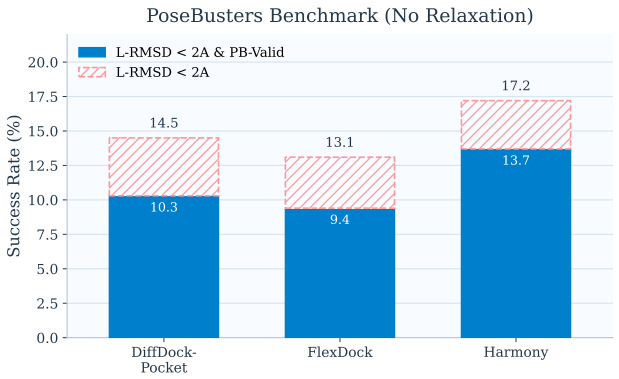


Figure 3. PoseBusters results with **no relaxation**.

5. Experiments

5.1. Experimental Setup and Results

Benchmarks. We evaluate *Harmony* using the widely established PDBBind benchmark Liu et al. (2017), relying on structures predicted by ESMFold Lin et al. (2022) to approximate the distribution of unbound protein conformations. Given the focus on flexible docking, the evaluation considers the accuracy of both ligand placement and binding pocket geometry. All of the reported metrics are for top-1 samples ranked by confidence and do not use oracle information. More details on datasets are in Appendix E.1.

To ensure a consistent comparison, predicted pocket structures are first aligned to their corresponding reference conformations. Following this alignment, heavy-atom RMSD is computed for both the ligand and the pocket atoms. Performance for ligand predictions is summarized using the median RMSD as well as the proportion of cases below a 2Å threshold. For pocket atoms, accuracy is reported as the fraction of predictions achieving an RMSD below 1Å. More details on evaluation metrics can be found in Appendix E.2. Additionally, we evaluate on the PoseBusters (Buttenschoen et al., 2024) set which includes a wide range of filters aimed to ensure the physical validity of generated poses. Exact stratification between different validity checks can be seen in Appendix B

Baselines. On the PDBBind benchmark, *Harmony* is evaluated against both classical and learning-based baselines. We include established search-based docking methods such as SMINA (Koes et al., 2013) and GNINA (McNutt et al., 2021), as well as recent flexible, pocket-level machine learning approaches including DiffDock-Pocket Plainer et al. (2023), ReDock Huang et al. (2024) and FlexDock Corso et al. (2025).

A notable consideration in these comparisons is the use of post-processing or relaxation steps. Several methods incorporate external energy minimization procedures which are often implemented via tools such as OpenMM or rely on auxiliary learned models to refine predicted complexes as in FlexDock. While such procedures can improve geometric plausibility by reducing steric clashes and energy artifacts, they may also obscure deficiencies in the underlying generative model by correcting both inter- and intramolecular inconsistencies.

To isolate the intrinsic performance of flexible docking models, we conduct additional comparisons on the PoseBusters benchmark using base versions of DiffDock-Pocket and recent FlexDock, explicitly excluding any relaxation or post-processing components. This allows for a more direct assessment of the generative models themselves, independent of downstream correction mechanisms.

Main Results. Table 1 summarizes the performance of prior approaches alongside *Harmony*, where consistent improvements are observed across multiple evaluation metrics. *Harmony* adopts the same architectural design and training protocol as DiffDock-Pocket and FlexDock. Under this setting, we observe gains in both ligand placement accuracy, measured by the fraction of predictions with RMSD < 2Å with the increase from 39.7% to 64.2%, and pocket reconstruction quality where AA-RMSD < 1Å increases from 41.7% to 47.3%.

An important distinction is that FlexDock relies on an additional relaxation module, which must be trained separately and applied as a post-processing step. In contrast, the model proposed here directly generates final complex structures without any auxiliary refinement, demonstrating that the improvements stem from the core generative framework rather than downstream correction procedures. On PoseBusters

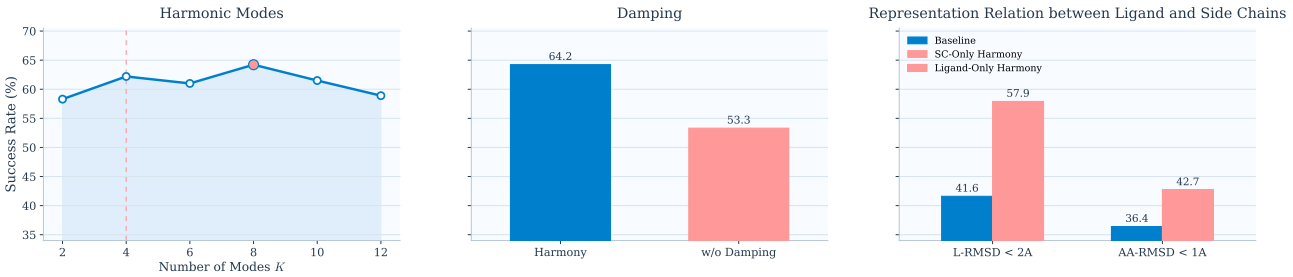


Figure 4. Ablations of *Harmony*. **Left**: Ablation of the number of harmonic modes K on PDBBind success rate. A smaller number of modes yields the best performance. **Middle**: Ablation of the VE heat-flow damping term. Removing damping degrades docking success, showing that explicit noise-dependent suppression of high-frequency torsional modes is important. **Right**: Applying harmonic learning only to ligand torsions improves side-chain modeling, while applying it only to side chains improves ligand success rate. This indicates that *Harmony* improves the joint representation used for flexible docking.

(Figure 3), our method outperforms direct pocket-based flexible competitors DiffDock-Pocket and FlexDock.

Overall, these results suggest that the Harmonic Torsional Diffusion framework introduces effective inductive biases for jointly modeling ligand placement and side-chain conformations, leading to improved structural accuracy.

5.2. Ablation Studies

Ablation Study on Harmonic Modes. We vary the number of harmonic modes K the model learns during training. If performance does not drop significantly from decreasing maximum K it would indicate coarse dynamics of torsional movement from apo to holo structures. This is indeed the case: in right part of Figure 4, we observe that *Harmony* performs the best with $K = 4$. Performance drops manageably for minimum $K = 2$. The model starts to lose performance further when too many harmonic modes are used. That can be explained by the above idea of intrinsic low-frequency nature. For all of K, *Harmony* outperforms latest rival FlexDock, which highlights the stability against hyperparameter choices.

Ablation Study on Frequency Damping. We believe *Harmony's* results are tied to the damping effect on torsional motion frequencies. In this ablation we remove the damping and allow model to learn all frequencies by itself. Results are reported in Figure 4 (middle) and conclude, that in such setting performance degrades and noising of high frequency components is the central part of the proposed model.

Ablation Study on Learning Signal Between Ligand and Side Chains. We would like to additionally answer, whether applying *Harmony's* framework simultaneously to ligand and side-chain torsions improves their performance. To validate this, we trained two versions of our model: *Harmony* only for ligand torsions and *Harmony* only for side-chain torsions. As a result, ligand success rate drops without harmonic model for side chains and vice versa (see right side of Figure 4). This supports the claim that representations learned by *Harmony* benefit both ligand and protein positions.

5.3. Case Studies

Beyond aggregate benchmarks, we examine two representative targets: EBNA1 (6NPP) (Messick et al., 2019) and KRAS G12D (6GJ8) (Kessler et al., 2019). For each complex, we compare the best confidence-ranked pose from the baseline model (FlexDock) and from *Harmony*. To verify that these targets are not unusually close to the training set, we additionally computed ligand, protein, and contact-based similarity scores to PDBBind training complexes as described in Appendix E.3. The mean combined similarity across the test set is 0.196 ± 0.019, while 6GJ8 and 6NPP score 0.189 and 0.183, respectively, indicating that both case studies are less similar than a typical test target. We emphasize that these similarity scores are only a coarse check against obvious train-test leakage and do not directly measure pocket difficulty, which depends primarily on the spatial arrangement and chemistry of local residues. In all cases, docking is performed from an apo receptor conformation, making the setting substantially more challenging than rigid docking.

EBNA1 (6NPP). EBNA1 is required for Epstein-Barr virus genome maintenance and is a therapeutically relevant viral target (Frappier, 2012; Jiang et al., 2018). Geometrically, the ligand binds in a broad site with a surrounding loop. The baseline prediction fails to preserve this local closure, whereas *Harmony* maintains the enclosing loop structure and improves both ligand and receptor pose.

KRAS G12D (6GJ8). KRAS is a central signaling GTPase and one of the most important oncogenic targets in cancer (Prior et al., 2012). The binding site in 6GJ8 is shallow and highly exposed, so the ligand is only weakly constrained by shape complementarity. This makes the complex

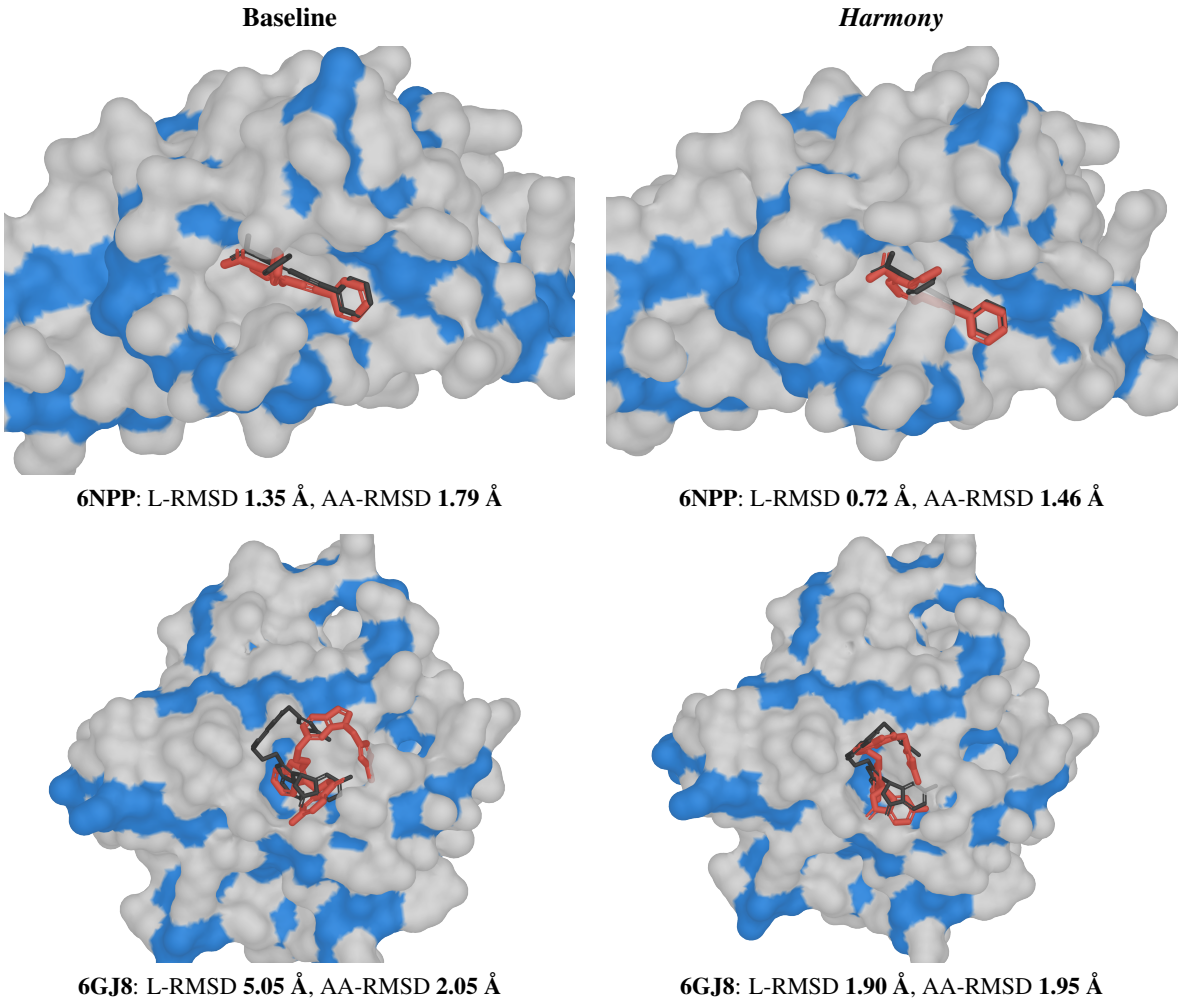


Figure 5. **Case studies on EBNA1 (6NPP) and KRAS G12D (6GJ8).** Ground-truth ligand is shown in **black**, and the predicted ligand is shown in **red**. The receptor is shown in its predicted configuration. In both examples, *Harmony* produces a more accurate ligand pose and improved local all-atom reconstruction than the baseline.

particularly difficult for docking from apo structure. In this regime, the baseline model produces a large ligand pose error, while *Harmony* recovers a substantially more accurate conformation closer to the crystal reference.

6. Discussion

Future Work and Limitations. The main limitation is that *Harmony* is based on a VE diffusion process that needs to have Gaussian noise prior. Thus, adapting this harmonic framework to a generative model for arbitrary endpoint distributions like flow-based interpolants is a compelling future direction as it would enable fully all-atom flexible docking with backbone motion. Additionally, it would be interesting to test harmonic score formulation in other tasks requiring generation of geometric graphs such as protein or material design.

Conclusion. We introduced *Harmony*, a harmonic torsional diffusion module for flexible protein-ligand docking. By parameterizing ligand and side-chain torsional scores in Fourier space and utilizing variance-exploding diffusion properties on the torus, *Harmony* makes angular periodicity explicit and improves the modeling of flexible binding sites. Across standard benchmarks and case studies, this simple geometric change consistently improves pose recovery and all-atom reconstruction. More broadly, our results suggest that respecting the intrinsic manifold structure of torsional motion is not a minor modeling detail, but a key ingredient for accurate diffusion-based flexible docking.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of

biomolecular interactions with alphafold 3. *Nature*, 630 (8016):493–500, 2024.

Alberts, I. L., Todorov, N. P., and Dean, P. M. Receptor flexibility in de novo ligand design and docking. *Journal of Medicinal Chemistry*, 48(21):6585–6596, 2005.

Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Bahar, I., Lezon, T. R., Yang, L.-W., and Eyal, E. Global dynamics of proteins: bridging between structure and function. *Annual review of biophysics*, 39:23–42, 2010.

Bharadwaj, V., Glover, A., Buluç, A., and Demmel, J. An efficient sparse kernel generator for o (3)-equivariant deep networks. In *2025 Proceedings of the Conference on Applied and Computational Discrete Algorithms (ACDA)*, pp. 32–46. SIAM, 2025.

Branden, C. I. and Tooze, J. *Introduction to protein structure*. Garland Science, 2012.

Buttenschoen, M., Morris, G. M., and Deane, C. M. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.

Chen, R. T. and Lipman, Y. Flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

Clark, J. J., Benson, M. L., Smith, R. D., and Carlson, H. A. Inherent versus induced protein flexibility: Comparisons within and between apo and holo structures. *PLoS computational biology*, 15(1):e1006705, 2019.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Corso, G., Somnath, V. R., Getz, N., Barzilay, R., Jaakkola, T., and Krause, A. Composing unbalanced flows for flexible docking and relaxation. In *The Thirteenth International Conference on Learning Representations*, 2025.

De Bortoli, V., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

Eastman, P., Galvelis, R., Peláez, R. P., Abreu, C. R., Farr, S. E., Gallicchio, E., Gorenko, A., Henry, M. M., Hu, F., Huang, J., et al. Openmm 8: molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1):109–116, 2023.

Frappier, L. The epstein-barr virus ebna1 protein. *Scientifica*, 2012(1):438204, 2012.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.

Gaudreault, F., Chartier, M., and Najmanovich, R. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics*, 28(18):i423–i430, 2012.

Geiger, M. and Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Huang, Y., Zhang, O., Wu, L., Tan, C., Lin, H., Gao, Z., Li, S., Li, S., et al. Re-dock: towards flexible and realistic molecular docking with diffusion bridge. *arXiv preprint arXiv:2402.11459*, 2024.

Jiang, L., Xie, C., Lung, H. L., Lo, K. W., Law, G.-L., Mak, N.-K., and Wong, K.-L. Ebna1-targeted inhibitors: Novel approaches for the treatment of epstein-barr virus-associated cancers. *Theranostics*, 8(19):5307, 2018.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *Advances in neural information processing systems*, 35: 24240–24253, 2022.

Kessler, D., Gmachl, M., Mantoulidis, A., Martin, L. J., Zoephel, A., Mayer, M., Gollner, A., Covini, D., Fischer, S., Gerstberger, T., et al. Drugging an undruggable pocket on kras. *Proceedings of the National Academy of Sciences*, 116(32):15823–15829, 2019.

Koes, D. R., Baumgartner, M. P., and Camacho, C. J. Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise. *Journal of chemical information and modeling*, 53(8):1893–1904, 2013.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv*, 2022:500902, 2022.

Liu, Z., Su, M., Han, L., Liu, J., Yang, Q., Li, Y., and Wang, R. Forging the basis for developing protein–ligand interaction scoring functions. *Accounts of chemical research*, 50(2):302–309, 2017.

Lu, W., Zhang, J., Huang, W., Zhang, Z., Jia, X., Wang, Z., Shi, L., Li, C., Wolynes, P. G., and Zheng, S. Dynamicbind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1):1071, 2024.

Mardia, K. V. and Jupp, P. E. *Directional Statistics*. John Wiley & Sons, 2009.

McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., and Koes, D. R. Gnina 1.0: molecular docking with deep learning. *Journal of cheminformatics*, 13(1):43, 2021.

Messick, T. E., Smith, G. R., Soldan, S. S., McDonnell, M. E., Deakyne, J. S., Malecka, K. A., Tolvinski, L., van den Heuvel, A. P. J., Gu, B.-W., Cassel, J. A., et al. Structure-based design of small-molecule inhibitors of ebna1 dna binding blocks epstein-barr virus latent infection and tumor growth. *Science translational medicine*, 11(482):eaau5612, 2019.

Miao, Z. and Cao, Y. Quantifying side-chain conformational variations in protein structure. *Scientific reports*, 6(1): 37024, 2016.

Morgan, H. L. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of chemical documentation*, 5(2):107–113, 1965.

Needleman, S. B. and Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.

Pagadala, N. S., Syed, K., and Tuszynski, J. Software for molecular docking: a review. *Biophysical reviews*, 9(2): 91–102, 2017.

Plainer, M., Toth, M., Dobers, S., Stark, H., Corso, G., Marquet, C., and Barzilay, R. Diffdock-pocket: Diffusion for pocket-level docking with sidechain flexibility. 2023.

Prat, A., Zhang, L., Deane, C. M., Teh, Y. W., and Morris, G. M. Sigmadock: Untwisting molecular docking with fragment-based se (3) diffusion. *arXiv preprint arXiv:2511.04854*, 2025.

Prior, I. A., Lewis, P. D., and Mattos, C. A comprehensive survey of ras mutations in cancer. *Cancer research*, 72 (10):2457–2467, 2012.

Sarkar, K. Score shocks: The burgers equation structure of diffusion generative models. *arXiv preprint arXiv:2604.07404*, 2026.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Stein, E. M. and Shakarchi, R. *Fourier Analysis: An Introduction*. Princeton University Press, 2003.

Teague, S. J. Implications of protein flexibility for drug discovery. *Nature reviews Drug discovery*, 2(7):527–541, 2003.

Thomsen, R. and Christensen, M. H. Moldock: a new technique for high-accuracy molecular docking. *Journal of medicinal chemistry*, 49(11):3315–3321, 2006.

Trott, O. and Olson, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Weikl, T. R. and Paul, F. Conformational selection in protein binding and function. *Protein Science*, 23(11):1508–1518, 2014.

A. Proofs of Propositions

Proof of Proposition 4.1. For the driftless SDE

$$d\mathbf{z}_\lambda = g(\lambda)\, d\mathbf{w}_\lambda,$$

the infinitesimal generator is

$$\mathcal{L}_\lambda = \frac{g(\lambda)^2}{2}\Delta.$$

Since the Laplacian is self-adjoint, the Fokker–Planck equation is

$$\partial_\lambda p_\lambda = \mathcal{L}_\lambda^* p_\lambda = \mathcal{L}_\lambda p_\lambda = \frac{g(\lambda)^2}{2}\Delta p_\lambda.$$

Thus the forward density evolves by a heat equation with time-dependent diffusivity $\nu(\lambda) = g(\lambda)^2/2$. □

Proof of Proposition 4.2. On the circle $\mathbb{T}$ with angular coordinate ϑ, the Laplace–Beltrami operator is $\Delta_\mathbb{T} = \partial_\vartheta^2$. Hence

$$\Delta_\mathbb{T} \cos(k\vartheta) = \partial_\vartheta^2 \cos(k\vartheta) = -k^2 \cos(k\vartheta), \qquad \Delta_\mathbb{T} \sin(k\vartheta) = \partial_\vartheta^2 \sin(k\vartheta) = -k^2 \sin(k\vartheta),$$

so $\cos(k\vartheta)$ and $\sin(k\vartheta)$ are eigenfunctions with eigenvalue $-k^2$. Therefore, for

$$\psi(\vartheta) = \sum_{k=1}^{K} \big[a_k \cos(k\vartheta) + b_k \sin(k\vartheta)\big],$$

the heat semigroup acts modewise:

$$\exp\left(\frac{\sigma^2}{2}\Delta_\mathbb{T}\right)\cos(k\vartheta) = \exp\left(-\frac{1}{2}k^2\sigma^2\right)\cos(k\vartheta), \qquad \exp\left(\frac{\sigma^2}{2}\Delta_\mathbb{T}\right)\sin(k\vartheta) = \exp\left(-\frac{1}{2}k^2\sigma^2\right)\sin(k\vartheta),$$

which yields

$$\exp\left(\frac{\sigma^2}{2}\Delta_\mathbb{T}\right)\psi(\vartheta) = \sum_{k=1}^{K} \exp\left(-\frac{1}{2}k^2\sigma^2\right)\big[a_k \cos(k\vartheta) + b_k \sin(k\vartheta)\big].$$

□

Proof of Proposition 4.3. Let

$$\vartheta_\lambda = \mathrm{wrap}(\vartheta_0 + \sigma(\lambda)\epsilon), \qquad \epsilon \sim \mathcal{N}(0,1).$$

Before wrapping, $\vartheta_0 + \sigma(\lambda)\epsilon$ is Gaussian with mean ϑ_0 and variance $\sigma^2(\lambda)$. Passing to the quotient $\mathbb{R}/2\pi\mathbb{Z} \cong \mathbb{T}$ periodizes this Gaussian, giving the wrapped normal kernel

$$p_\lambda(\vartheta \mid \vartheta_0) = \sum_{n\in\mathbb{Z}} \frac{1}{\sqrt{2\pi\sigma^2(\lambda)}} \exp\left(-\frac{(\vartheta - \vartheta_0 - 2\pi n)^2}{2\sigma^2(\lambda)}\right).$$

The heat kernel on $\mathbb{T}$ at time t is exactly the periodized Gaussian on $\mathbb{R}$ with variance $2t$, so setting $t = \sigma^2(\lambda)/2$ gives

$$p_\lambda(\cdot \mid \vartheta_0) = \exp\left(\frac{\sigma^2(\lambda)}{2}\Delta_\mathbb{T}\right)\delta_{\vartheta_0}.$$

Differentiating with respect to λ yields

$$\partial_\lambda p_\lambda = \frac{(\sigma^2)'(\lambda)}{2}\Delta_\mathbb{T}\, p_\lambda.$$

In particular, if $\sigma^2(\lambda) = \int_0^\lambda g(s)^2\, ds$, then $(\sigma^2)'(\lambda) = g(\lambda)^2$, and therefore

$$\partial_\lambda p_\lambda = \frac{g(\lambda)^2}{2}\Delta_\mathbb{T}\, p_\lambda.$$

□

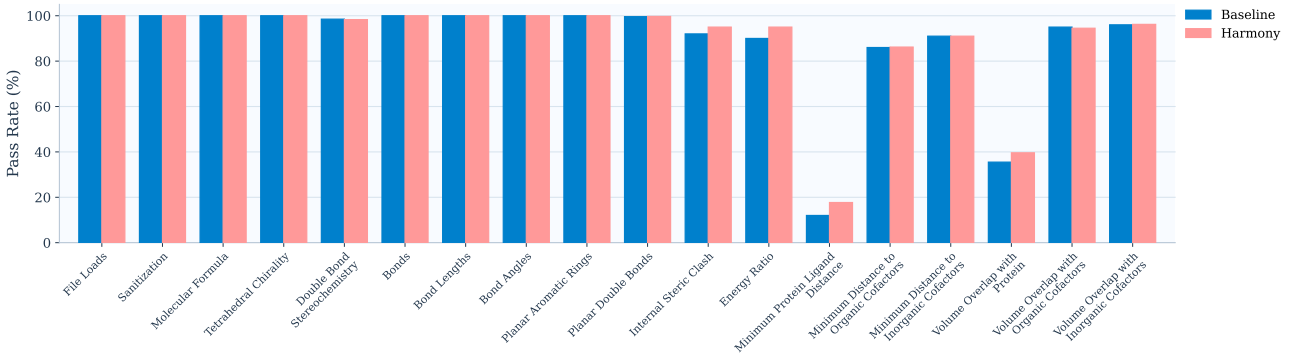


Figure 6. PoseBusters check-wise breakdown for the baseline model and *Harmony*. Each bar reports the pass rate for one PoseBusters criterion, highlighting that *Harmony* improves several physically meaningful checks, particularly those related to internal steric validity, ligand strain, and protein–ligand separation.

B. Stratification of Validity Checks in PoseBusters

In this section we provide a detailed breakdown of validity check passes in PoseBusters for baseline model (FlexDock) and *Harmony* with no energy minimization. In Figure 6 it can be seen that our framework improves on protein-ligand distance, internal steric clashes and ligand strain.

C. Training and Inference

In this section we provide additional details on training and inference algorithms for our model, and data preparation such as pocket selection, conformer matching, alignment of protein frames and flexibility formalization.

C.1. Algorithms for Training and Inference of *Harmony*

We provide comprehensive pseudocode for training and inference of *Harmony* in Algorithm 1 and Algorithm 2.

C.2. Data Preparation

Our data preparation largely follows DiffDock-Pocket (Plainer et al., 2023) and FlexDock (Corso et al., 2025). Below are more details.

Pocket Selection. We use pocket-centered docking to reduce the search space and focus the model on the local receptor environment most relevant for binding. Starting from the all-atom apo receptor and the reference ligand pose, we first identify a set of nearby residues by computing the minimum distance from each receptor residue to the ligand and selecting residues within a pocket cutoff. In the implementation, this is done at the atom level and then aggregated to the residue level by taking the minimum atom–ligand distance within each residue. Let S_{near} denote this initial set of nearby residues. We then define the pocket center as the mean of the corresponding apo C_α coordinates,

$$\mathbf{c}_{\text{pocket}} = \frac{1}{|S_{\text{near}}|} \sum_{r \in S_{\text{near}}} \mathbf{x}^{\text{apo}}_{r,C_\alpha}.$$

We additionally enlarge the crop by retaining all residues whose apo C_α lies within a fixed buffer radius of $\mathbf{c}_{\text{pocket}}$ which is chosen to be 10 Å. The buffer is needed to make model for robust to the choice of a pocket as noted in DiffDock-Pocket (Plainer et al., 2023) and FlexDock (Corso et al., 2025).This yields a pocket-centered residue subset together with the corresponding protein atoms. If the initial cutoff would produce too few residues, the nearest residues are added until a minimum pocket size is reached. After cropping all coordinates are recentered by subtracting $\mathbf{c}_{\text{pocket}}$. At inference time, the holo receptor coordinates are not available and we define pocket center for the apo receptor.

Conformer Matching for Ligand and Side Chains. Crystal structures are used as targets, and their distance variables such as bond lengths or angles fundamentally differ from ones, produced by generated ligand conformers (e.g. in RDKit)

Algorithm 1 Training of *Harmony*

Require: Preprocessed training set $\mathcal{D} = \{(\mathcal{G}^{(i)}_{\text{pocket}}, \mathbf{z}_0^{(i)})\}_{i=1}^{M}$, where $\mathbf{z}_0 = (\mathbf{t}_0, \mathbf{R}_0, \boldsymbol{\tau}_0, \boldsymbol{\chi}_0)$; noise schedules $\sigma_{\text{tr}}, \sigma_{\text{rot}}, \sigma_{\text{tor}}, \sigma_{\text{sc}}$; model f_θ

Ensure: Updated parameters θ

for all minibatches $\mathcal{B} \subset \mathcal{D}$ **do**

Sample $\lambda \sim \mathcal{U}(0, 1)$

Sample translation noise $\boldsymbol{\epsilon}_{\text{tr}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$

Sample rotational noise $\boldsymbol{\omega}_\lambda \sim \text{IGSO}(3; \sigma_{\text{rot}}(\lambda))$

Sample torsional noises $\boldsymbol{\epsilon}_{\text{tor}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_l})$ and $\boldsymbol{\epsilon}_{\text{sc}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_s})$

Construct the noised docking state:

$$\mathbf{t}_\lambda = \mathbf{t}_0 + \sigma_{\text{tr}}(\lambda)\boldsymbol{\epsilon}_{\text{tr}}, \qquad \mathbf{R}_\lambda = \text{Exp}(\boldsymbol{\omega}_\lambda)\mathbf{R}_0,$$

$$\boldsymbol{\tau}_\lambda = \text{wrap}(\boldsymbol{\tau}_0 + \sigma_{\text{tor}}(\lambda)\boldsymbol{\epsilon}_{\text{tor}}), \qquad \boldsymbol{\chi}_\lambda = \text{wrap}(\boldsymbol{\chi}_0 + \sigma_{\text{sc}}(\lambda)\boldsymbol{\epsilon}_{\text{sc}}).$$

Apply $(\mathbf{t}_\lambda, \mathbf{R}_\lambda, \boldsymbol{\tau}_\lambda, \boldsymbol{\chi}_\lambda)$ to the clean pocket graph to obtain $\mathcal{G}_\lambda$

Predict rigid-body scores and harmonic coefficients:

$$\hat{\mathbf{s}}_{\text{tr}},\ \hat{\mathbf{s}}_{\text{rot}},\ \{\hat{a}^{\,l}_{e,k}, \hat{b}^{\,l}_{e,k}\},\ \{\hat{a}^{\,sc}_{e,k}, \hat{b}^{\,sc}_{e,k}\} \leftarrow f_\theta(\mathcal{G}_\lambda, \lambda).$$

Evaluate ligand and side-chain torsional scores from the harmonic heads:

$$\hat{\mathbf{s}}_{\text{tor}} \leftarrow \partial_\vartheta \hat{\psi}^{\,l}_\lambda(\vartheta), \qquad \hat{\mathbf{s}}_{\text{sc}} \leftarrow \partial_\vartheta \hat{\psi}^{\,sc}_\lambda(\vartheta),$$

where each $\hat{\psi}_{e,\lambda}$ is the heat-damped harmonic potential from Equation (1).

Compute score targets from the known forward kernel:

$$\mathbf{s}^\star_{\text{tr}},\ \mathbf{s}^\star_{\text{rot}},\ \mathbf{s}^\star_{\text{tor}},\ \mathbf{s}^\star_{\text{sc}} \leftarrow \nabla \log q_\lambda(\mathbf{z}_\lambda \mid \mathbf{z}_0).$$

Compute the training loss:

$$\mathcal{L} = \mathcal{L}_{\text{tr}} + \mathcal{L}_{\text{rot}} + \mathcal{L}_{\text{tor}} + \mathcal{L}_{\text{sc}},$$

with each term matching the corresponding predicted and target scores.

Update θ with Adam on $\mathcal{L}$

end for

or apo receptor poses derived from folding models (e.g ESMFold). Thus, usage of crystal coordinates would induce distributional shift during inference. As discussed in Plainer et al. (2023) and Corso et al. (2025), to tackle this problem we need to perform conformer matching for ligand and side chains before training. Full algorithm can be examined in Algorithm 4.

For ligands, we generate a fresh conformer using RDKit with a force field refinement, identify the rotatable bonds, and then optimize the ligand torsion angles by differential evolution so that the generated conformer best matches the reference ligand pose. The resulting matched conformer is used as the input ligand, while the original ligand coordinates are retained as the reference pose.

For receptor side chains, after apo-holo alignment, we optimize each flexible residue independently over its torsional degrees of freedom using differential evolution, with the objective of minimizing the RMSD between the moved side-chain atoms and the corresponding holo coordinates. The optimized torsional differences are stored per rotatable side-chain bond, and applying them to the aligned apo coordinates yields a conformer-matched apo structure whose side chains are as close as possible to the holo arrangement. This matched structure provides the target for our model.

Graph Flexibility. Flexibility is represented through directed torsional edges and binary masks indicating which bonds are actually rotatable. General algorithm is demonstrated in Algorithm 3. For ligands, a covalent bond is considered rotatable if removing the corresponding undirected bond disconnects the molecular graph and the moved fragment contains more than one atom. Among the two directed versions of the bond, we retain the direction associated with the fragment that will be

Algorithm 2 Inference with *Harmony*

Require: Pocket-centered apo graph $\mathcal{G}^{\text{apo}}_{\text{pocket}}$, initial ligand conformer, number of reverse steps N, noise schedules $\sigma_{\text{tr}}, \sigma_{\text{rot}}, \sigma_{\text{tor}}, \sigma_{\text{sc}}$, trained model f_θ

Ensure: Predicted bound ligand pose and pocket side-chain arrangement

1: Initialize:

$$\mathbf{t}_N \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{tr,max}}\mathbf{I}_3), \qquad \mathbf{R}_N \sim \mathcal{U}(SO(3)),$$
$$\boldsymbol{\tau}_N \sim \mathcal{U}(\mathbb{T}^{m_l}), \qquad \boldsymbol{\chi}_N \sim \mathcal{U}(\mathbb{T}^{m_s}).$$

2: Apply $(\mathbf{t}_N, \mathbf{R}_N, \boldsymbol{\tau}_N, \boldsymbol{\chi}_N)$ to the initial ligand and apo pocket side chains to obtain $\mathcal{G}_N$

3: **for** $n = N, N-1, \ldots, 1$ **do**

4: Set $\lambda_n = n/N$, $\lambda_{n-1} = (n-1)/N$, and $\Delta\lambda = \lambda_n - \lambda_{n-1}$

5: Predict rigid-body scores and harmonic coefficients:

$$\hat{\mathbf{s}}_{\text{tr}},\ \hat{\mathbf{s}}_{\text{rot}},\ \{\hat{a}^{\,l}_{e,k}, \hat{b}^{\,l}_{e,k}\},\ \{\hat{a}^{\,sc}_{e,k}, \hat{b}^{\,sc}_{e,k}\} \leftarrow f_\theta(\mathcal{G}_n, \lambda_n).$$

6: Evaluate torsional scores from the harmonic heads:

$$\hat{\mathbf{s}}_{\text{tor}} \leftarrow \partial_\vartheta \hat{\psi}^{\,l}_{\lambda_n}(\vartheta), \qquad \hat{\mathbf{s}}_{\text{sc}} \leftarrow \partial_\vartheta \hat{\psi}^{\,sc}_{\lambda_n}(\vartheta).$$

7: Sample reverse-time noises:

$$\boldsymbol{\xi}_{\text{tr}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3), \quad \boldsymbol{\xi}_{\text{rot}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3),$$
$$\boldsymbol{\xi}_{\text{tor}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_l}), \quad \boldsymbol{\xi}_{\text{sc}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_s}).$$

8: Perform one reverse VE update:

$$\mathbf{t}_{n-1} = \mathbf{t}_n + g_{\text{tr}}(\lambda_n)^2 \Delta\lambda\, \hat{\mathbf{s}}_{\text{tr}} + g_{\text{tr}}(\lambda_n)\sqrt{\Delta\lambda}\, \boldsymbol{\xi}_{\text{tr}},$$

$$\mathbf{R}_{n-1} = \text{Exp}\Big(g_{\text{rot}}(\lambda_n)^2 \Delta\lambda\, \hat{\mathbf{s}}_{\text{rot}} + g_{\text{rot}}(\lambda_n)\sqrt{\Delta\lambda}\, \boldsymbol{\xi}_{\text{rot}} \Big)\, \mathbf{R}_n,$$

$$\boldsymbol{\tau}_{n-1} = \text{wrap}\Big(\boldsymbol{\tau}_n + g_{\text{tor}}(\lambda_n)^2 \Delta\lambda\, \hat{\mathbf{s}}_{\text{tor}} + g_{\text{tor}}(\lambda_n)\sqrt{\Delta\lambda}\, \boldsymbol{\xi}_{\text{tor}} \Big),$$

$$\boldsymbol{\chi}_{n-1} = \text{wrap}\Big(\boldsymbol{\chi}_n + g_{\text{sc}}(\lambda_n)^2 \Delta\lambda\, \hat{\mathbf{s}}_{\text{sc}} + g_{\text{sc}}(\lambda_n)\sqrt{\Delta\lambda}\, \boldsymbol{\xi}_{\text{sc}} \Big).$$

9: Apply $(\mathbf{t}_{n-1}, \mathbf{R}_{n-1}, \boldsymbol{\tau}_{n-1}, \boldsymbol{\chi}_{n-1})$ to obtain the updated graph $\mathcal{G}_{n-1}$

10: **end for**

11: **return** final ligand coordinates and pocket side-chain coordinates induced by $(\mathbf{t}_0, \mathbf{R}_0, \boldsymbol{\tau}_0, \boldsymbol{\chi}_0)$

rotated and store the list of affected atoms for each torsion. This induces a mask over ligand bond edges together with a fragment index that specifies which atoms move when a given torsion is perturbed.

For receptor side chains, we construct a directed side-chain bond graph residue by residue, starting from standard PDB atom names and moving outward from C_α along the side-chain topology. Glycine has no side-chain torsions and contributes no rotatable edges. Proline side-chain bonds are included in the graph but are marked non-rotatable. For all other residues, the same graph criterion is applied to identify side-chain bonds whose rotation moves a particular fragment.

Alignment of Apo-Holo Frames. The purpose of apo-holo alignment is to disentangle rigid-body differences, local backbone frame differences, and genuine side-chain conformational changes before defining the diffusion targets. The first step is a global Kabsch alignment of the apo receptor to the holo receptor using the C_α atoms. This removes the trivial rigid-body offset between the two structures. The second step aligns apo side chains into the local backbone frames of the holo structure. Concretely, for each residue we construct a local frame from the N-C_α-C backbone triplet, compute the rotation that maps the apo frame to the holo frame, and apply that rotation to the corresponding local side-chain coordinates. The resulting coordinates are stored as the aligned apo structure. At inference time, no holo structure is available, thus the apo structure itself defines the working frame without any alignment.

Algorithm 3 Finding Flexible Bonds and Rotated Fragments

Require: Directed bond graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, optional anchor atom v_{base}
Ensure: Rotatable-edge mask m_{rot}, fragment index $\mathcal{F}$

```
Initialize m_rot ← 0 and F ← ∅
for all undirected bonds {u, v} represented by directed edges (u, v) and (v, u) do
    Remove {u, v} from the undirected version of G
    if the graph remains connected then
        Mark both directions as non-rotatable
        continue
    end if
    Let C_1, C_2 be the two connected components
    if v_base is specified then
        Choose as rotated fragment the component that does not contain v_base
    else
        Choose as rotated fragment the smaller component
    end if
    Let C_rot be the chosen fragment and let e_rot be the directed edge whose orientation points into that fragment
    if |C_rot| > 1 then
        Set m_rot(e_rot) ← 1
    end if
    Store the rotated atoms:
        F ← F ∪ {(e_rot, a) : a ∈ C_rot}
end for
Construct a dependency graph between rotatable edges from fragment overlap and topologically sort them
Reorder F according to this torsional application order
return m_rot, F
```

D. Model Architecture

D.1. Graph Construction.

Harmony uses a heterogeneous representation of the pocket complex with three node sets: ligand atoms, receptor residues at C_α positions, and receptor atoms. Following FlexDock (Corso et al., 2025), we distinguish between *static* structural graphs constructed during preprocessing and *dynamic* geometric graphs recomputed online from the current noised coordinates during message passing.

Static Structural Graphs. The ligand is first represented by its bidirected covalent bond graph extracted from the RDKit molecular graph. Each covalent bond contributes two directed edges, and the corresponding bond type is stored as a one-hot edge feature over the bond classes $\{\text{single}, \text{double}, \text{triple}, \text{aromatic}, \text{dative}\}$. This covalent graph is also used to define ligand torsional degrees of freedom through the rotatable edge masks described in Appendix C.2.

For the receptor, we store one residue node per amino acid at the C_α coordinate, and one atom node per protein atom in the selected pocket. We additionally store a deterministic atom-to-residue mapping edge connecting each receptor atom to its parent residue. Side-chain torsional topology is cached separately as a directed residue-wise bond graph constructed from standard PDB atom naming.

Dynamic Geometric Graphs. During training and inference, the equivariant backbone rebuilds its spatial neighborhoods from the current noised coordinates. The ligand atom graph is augmented with a geometric radius graph using a 6 Å cutoff. Thus, ligand message passing is performed on the union of covalent edges and local geometric edges.

The residue-level receptor graph is constructed with a k-nearest-neighbor graph on C_α coordinates, using at most 24 neighbors per residue, followed by a radius filter at 15 Å. The receptor atom graph is built analogously with at most 12 neighbors per atom and a radius cutoff of 6 Å.

Algorithm 4 Conformer Matching for Ligands and Side Chains

Require: Reference ligand pose $\mathbf{X}^{l,\star}$, apo receptor coordinates $\mathbf{X}^a$, holo receptor coordinates $\mathbf{X}^h$
Ensure: Matched ligand conformer $\tilde{\mathbf{X}}^l$, matched apo side-chain coordinates $\tilde{\mathbf{X}}^a$, torsional updates $\Delta\boldsymbol{\chi}$

Ligand conformer matching:
Generate a fresh ligand conformer $\mathbf{X}^{l,0}$ with RDKit and force-field refinement
Identify ligand rotatable bonds $\mathcal{B}^l$
Optimize torsion angles $\boldsymbol{\tau}$ with differential evolution:

$$\hat{\boldsymbol{\tau}} = \arg \min_{\boldsymbol{\tau} \in [-\pi,\pi]^{|\mathcal{B}^l|}} \mathrm{RMSD}\big(\mathrm{Rotate}(\mathbf{X}^{l,0}, \boldsymbol{\tau}), \mathbf{X}^{l,\star}\big)$$

Set

$$\tilde{\mathbf{X}}^l = \mathrm{Rotate}(\mathbf{X}^{l,0}, \hat{\boldsymbol{\tau}})$$

Side-chain conformer matching:
Compute global apo–holo Kabsch alignment using receptor C_α atoms
Align apo side chains into local holo backbone frames using per-residue N–C_α–C frames
Initialize $\tilde{\mathbf{X}}^a \leftarrow \mathbf{X}^a_{\text{aligned}}$
for all flexible residues r **do**
  Extract residue-specific rotatable side-chain bonds $\mathcal{B}^a_r$
  if $\mathcal{B}^a_r = \emptyset$ **then**
    continue
  end if
  Optimize residue torsions independently with differential evolution:

$$\hat{\boldsymbol{\chi}}_r = \arg \min_{\boldsymbol{\chi}_r \in [-\pi,\pi]^{|\mathcal{B}^a_r|}} \mathrm{RMSD}\Big(\mathrm{RotateResidue}(\tilde{\mathbf{X}}^a, r, \boldsymbol{\chi}_r), \mathbf{X}^h\Big)$$

  if the optimized residue RMSD improves **then**
    Apply $\hat{\boldsymbol{\chi}}_r$ to $\tilde{\mathbf{X}}^a$
    Store the corresponding per-bond torsional updates in $\Delta\boldsymbol{\chi}$
  end if
end for
return $\tilde{\mathbf{X}}^l, \tilde{\mathbf{X}}^a, \Delta\boldsymbol{\chi}$

We use three types of cross-graph interactions. First, ligand atoms are connected to receptor residue nodes by a radius graph with maximum distance 80 Å. Second, ligand atoms are connected to receptor atoms within 6 Å. Third, each receptor atom is connected to its parent residue through the static atom-to-residue assignment edges described above.

D.2. Featurization

Ligand Features. Ligand atom features follow the standard RDKit-based featurization used in DiffDock-Pocket and FlexDock. For each ligand atom, we encode atomic number, chirality, degree, formal charge, implicit valence, total hydrogen count, number of radical electrons, hybridization state, aromaticity, number of rings, and binary membership indicators for rings of sizes 3 through 8. Covalent ligand edges carry a one-hot bond-type encoding as described above. For non-covalent geometric ligand edges added online by the radius graph, the bond-type channel is set to zero.

Receptor Features. Each receptor residue node is featurized by amino-acid identity, and, a precomputed ESM-2 embedding is concatenated to this categorical residue representation. Each receptor atom node is featurized by four categorical attributes: the identity of the parent residue, the atomic number, a coarse atom-name category, and a fine atom-name category. The coarse atom-name category is obtained from the first two characters of the PDB atom name (e.g., `CA`, `CB`, `ND`), while the fine category uses the full atom name (e.g., `CA`, `CD1`, `NE2`).

In addition to these categorical features, we store several auxiliary atom-level quantities used by the model, including van der Waals radii and boolean masks identifying C_α, C, and N atoms. Each residue node is additionally augmented with the

two local backbone vectors $\mathbf{x}_N - \mathbf{x}_{C_\alpha}$ and $\mathbf{x}_C - \mathbf{x}_{C_\alpha}$, yielding a 6-dimensional continuous orientation feature.

Geometric and Diffusion Features. Beyond the static node features above, Harmony augments all node types with diffusion-time embeddings computed from the current noise level. We use sinusoidal time embeddings of dimension 32. Edge distances are expanded with radial basis functions of dimension 32 for both intra-graph and cross-graph interactions. The equivariant backbone further computes spherical harmonics from the relative edge vectors, so each geometric edge is represented by its source-node time embedding, distance expansion, and directional information derived from the relative coordinates.

Torsional Metadata. Finally, preprocessing stores the torsional metadata required for ligand and side-chain updates: rotatable edge masks, fragment indices specifying which atoms move under each torsion, and matched torsional targets obtained from conformer matching (Appendix C.2).

D.3. Model

Architecture Details. *Harmony* is instantiated as a heterogeneous e3nn (Geiger & Smidt, 2022) $SE(3)$-equivariant network over three coupled graphs: the ligand atom graph, the residue-level receptor backbone graph defined on C_α atoms, and the receptor atom graph in the selected pocket. The shared backbone uses 6 equivariant tensor-product convolution layers with hidden widths $n_s = 60$ for scalar channels and $n_v = 15$ for higher-order channels, and spherical harmonics up to degree $l_{\max} = 2$. We use layer normalization, SiLU activations, and dropout with rate 0.1 throughout.

Node features are first embedded from categorical atom-level and residue-level descriptors together with sinusoidal diffusion-time embeddings of dimension 32. Pairwise distances are expanded with RBFs of dimension 32 for both intra-graph and cross-graph edges. The equivariant message-passing neighborhoods are constructed separately at each scale, with ligand, residue, atom, and cross-graph interactions handled by the shared heterogeneous backbone. Full graph construction and featurization details are deferred to Section D.1.

On top of the shared equivariant trunk, *Harmony* uses three prediction modules. A rigid-body head predicts ligand translation and rotation scores. A ligand torsion head predicts harmonic coefficients for ligand torsional score fields, and a side-chain torsion head predicts harmonic coefficients for flexible pocket side-chain torsions. The side-chain torsion head additionally conditions the harmonic coefficients on the diffusion noise level. These predicted coefficients are converted into periodic torsional scores through the harmonic parameterization introduced in Section 4.3.

Confidence Head. Rather than training a separate confidence model, we attach a lightweight confidence head to the shared docking encoder. This head takes the pooled complex representation produced by the main docking network and predicts a scalar confidence score for each sampled pose, so confidence estimation is learned jointly with docking while adding only minimal computational overhead.

As the supervision target, we use protein–ligand interface lDDT (PLI-lDDT) (Lu et al., 2024), which measures how well the predicted complex preserves native receptor–ligand contact distances. Specifically, PLI-lDDT averages, over atom pairs that are in contact in the reference complex, the fraction of predicted distances that fall within several tolerance thresholds. This makes it a natural confidence target, since it reflects local binding-site quality rather than only global coordinate error.

To focus confidence learning on geometries that are close to the final denoised samples, we use a time-weighted regression loss. If $\hat{c}_n$ denotes the predicted confidence for sample n, c_n the corresponding PLI-lDDT target, and λ_n its diffusion level, we optimize

$$\mathcal{L}_{\text{conf}} = \frac{\sum_n w(\lambda_n)\,|\hat{c}_n - c_n|}{\sum_n w(\lambda_n)}, \qquad w(\lambda) = (1-\lambda)^2.$$

This weighting emphasizes low-noise states, which are more relevant to the final docking prediction, while reducing the influence of highly corrupted intermediate samples.

OpenEquivariance CUDA Kernels for Tensor Products. Our implementation uses OpenEquivariance (Bharadwaj et al., 2025) as the backend for the Clebsch–Gordan tensor products inside the e3nn equivariant convolutions. These tensor products are the main low-level bottleneck of $SE(3)$-equivariant message passing, since they are evaluated repeatedly across ligand, residue, and atom graphs at every layer. OpenEquivariance replaces generic tensor-product execution with JIT-compiled CUDA kernels specialized to the irreducible representation layout of the model. In particular, the generated

kernels exploit the structured sparsity of Clebsch-Gordan coefficient tensors, precompute efficient execution schedules from the fixed irrep structure, and reduce memory consumption by staging small contractions. In *Harmony*, enabling these kernels leaves the architecture and training objective unchanged, but substantially improves throughput of the equivariant backbone. Empirically, this reduces end-to-end training time from roughly 4.5 days to 2 days on 8 H100 80GB GPUs.

E. Experimental Details

E.1. Data

We train *Harmony* on PDBBind (Liu et al., 2017), using essentially the same data pipeline asFlexDock (Corso et al., 2025). In particular, we adopt the standard time-based split, where complexes deposited before 2019 are divided into training and validation sets, while 363 complexes deposited after 2019 are reserved for testing.

To construct paired holo and apo receptor structures, the protein-ligand complexes in PDBBind are processed them with PDBFixer from the OpenMM (Eastman et al., 2023) to standardize non-canonical residues and complete missing heavy atoms. These corrected complex structures define our holo references. We then extract the protein sequence from the processed receptor and predict an apo structure with ESMFold (Lin et al., 2022). The resulting repaired ESMFold structures are used as apo receptors. Hydrogen atoms are removed from both apo and holo structures.

In addition to pose-accuracy metrics, we evaluate the physical plausibility of generated complexes with PoseBusters (Buttenschoen et al., 2024). This benchmark assesses whether predicted poses satisfy basic chemical and geometric constraints, including ligand internal validity and protein-ligand interaction checks such as steric clash detection.

E.2. Evaluation Metrics

We evaluate docking quality using both ligand-level and complex-level geometric criteria. Our primary ligand metric is the ligand root-mean-square deviation (L-RMSD) between the predicted ligand pose and the reference crystal pose after optimal rigid alignment of the pockets. For a ligand with N_l atoms, this is

$$\text{L-RMSD} = \sqrt{\frac{1}{N_l}\sum_{i=1}^{N_l}\left\|\hat{\mathbf{x}}_i^l - \mathbf{x}_i^{l,\star}\right\|_2^2},$$

where $\hat{\mathbf{x}}_i^l$ and $\mathbf{x}_i^{l,\star}$ denote the predicted and reference coordinates of ligand atom i, respectively. Following standard docking practice, we report the fraction of complexes satisfying

$$\text{L-RMSD} < 2\text{Å},$$

which we refer to as the *ligand success rate*.

To assess whether the full predicted complex is geometrically accurate, including the flexible binding-site side chains, we also measure all-atom RMSD (AA-RMSD) over the receptor atoms retained in the pocket representation. Let N_a denote the total number of evaluated atoms. Then

$$\text{AA-RMSD} = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}\|\hat{\mathbf{x}}_i - \mathbf{x}_i^\star\|_2^2}.$$

We report the fraction of predictions with

$$\text{AA-RMSD} < 1\text{Å},$$

which gives a stricter measure of local all-atom reconstruction quality.

E.3. Similarity Analysis for Case Studies

To contextualize the qualitative case studies, we compared each target complex against the PDBBind training split using a simple descriptor-based similarity analysis.

Ligand Similarity. For each ligand, we compute a Morgan fingerprint (Morgan, 1965) with radius 2 and 1024 bits after removing hydrogens. Let $F_l(A)$ and $F_l(B)$ denote the sets of active fingerprint bits for two ligands A and B. Ligand similarity is measured by the Dice coefficient

$$S_{\text{lig}}(A, B) = \frac{2\,|F_l(A) \cap F_l(B)|}{|F_l(A)| + |F_l(B)|}.$$

Protein Similarity. For each receptor, we parse the protein structure file and convert the amino-acid sequence into a one-letter sequence. Protein similarity is then computed as global sequence identity obtained from a Needleman–Wunsch alignment (Needleman & Wunsch, 1970). If $L_{\text{match}}(A, B)$ denotes the number of matched residues in the optimal global alignment of proteins A and B, and $L_{\text{align}}(A, B)$ is the aligned length including gaps, then

$$S_{\text{prot}}(A, B) = \frac{L_{\text{match}}(A, B)}{L_{\text{align}}(A, B)}.$$

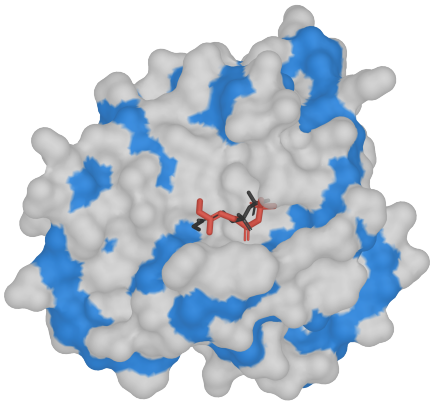
Contact Interaction Similarity. To obtain a coarse description of the local binding environment without relying on explicit interaction typing, we build a residue-contact signature from heavy-atom distances between the ligand and the receptor. For each receptor residue, we compute the minimum heavy-atom distance to the ligand and record the residue if that distance is below 4.5 Å. Contacts are further divided into two distance buckets: *close* for distances at most 3.5 Å and *near* for distances between 3.5 and 4.5 Å. Repeated occurrences of the same residue type within a bucket are counted separately, yielding a contact signature F_c. We then compare two complexes with the Jaccard similarity

$$S_{\text{cont}}(A, B) = \frac{|F_c(A) \cap F_c(B)|}{|F_c(A) \cup F_c(B)|}.$$

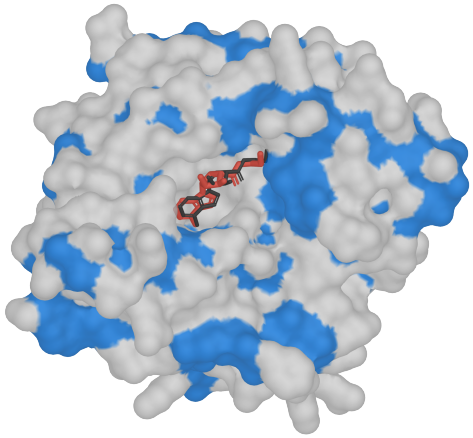
Combined Similarity Score. For each target complex, we compare its descriptor triplet against all complexes in the training split and average the three similarity components:

$$S_{\text{comb}}(A, B) = \frac{S_{\text{lig}}(A, B) + S_{\text{prot}}(A, B) + S_{\text{cont}}(A, B)}{3}.$$

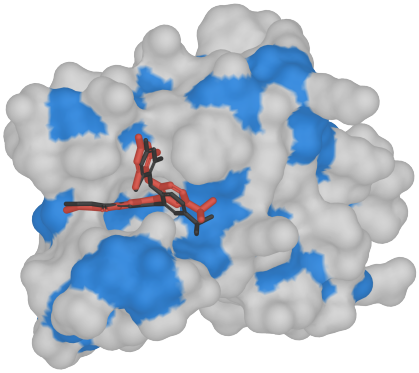
F. Visualizations



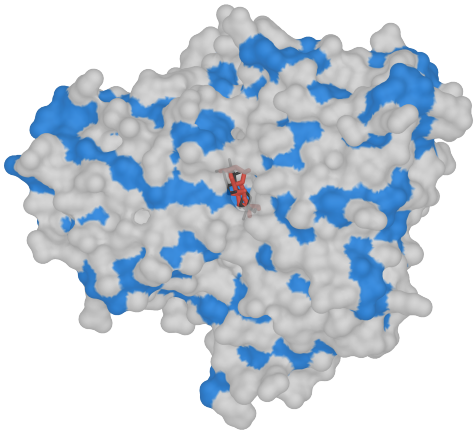
5ZXK: L-RMSD **2.53 Å**, AA-RMSD **1.19 Å**



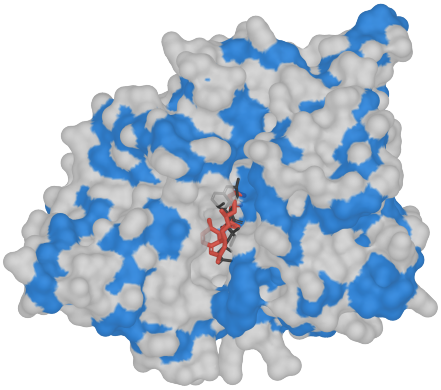
6CYG: L-RMSD **0.53 Å**, AA-RMSD **1.47 Å**



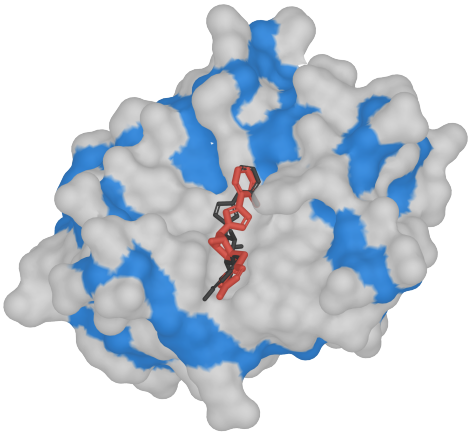
6E6J: L-RMSD **0.64 Å**, AA-RMSD **0.77 Å**



6JAN: L-RMSD **1.26 Å**, AA-RMSD **0.90 Å**



6KJD: L-RMSD **3.97 Å**, AA-RMSD **3.04 Å**



6QLT: L-RMSD **1.44 Å**, AA-RMSD **1.00 Å**

Figure 7. Extended results.