# Re-Dock: Towards Flexible and Realistic Molecular Docking with Diffusion Bridge

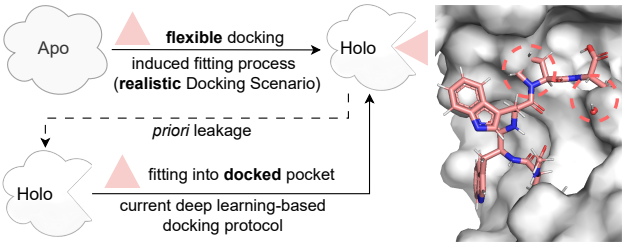
**Yufei Huang** [*1 2] **Odin Zhang** [*1 3] **Lirong Wu** [1 2] **Cheng Tan** [1 2] **Haitao Lin** [1 2]
**Zhangyang Gao** [1 2] **Siyuan Li** [1 2] **Stan. Z. Li** [† 2]

## Abstract

Accurate prediction of protein-ligand binding structures, a task known as molecular docking is crucial for drug design but remains challenging. While deep learning has shown promise, existing methods often depend on holo-protein structures (*docked*, and not accessible in *realistic* tasks) or neglect pocket sidechain conformations, leading to limited practical utility and unrealistic conformation predictions. To fill these gaps, we introduce an under-explored task, named *flexible docking* to predict poses of ligand and pocket sidechains simultaneously and introduce `Re-Dock`, a novel diffusion bridge generative model extended to geometric manifolds. Specifically, we propose energy-to-geometry mapping inspired by the Newton-Euler equation to co-model the binding energy and conformations for reflecting the energy-constrained docking generative process. Comprehensive experiments on designed benchmark datasets including apo-dock and cross-dock demonstrate our model's superior effectiveness and efficiency over current methods.

## 1. Introduction

Proteins can have their biological functions (Huang et al., 2023b;a; Wu et al., 2022b) altered by the binding of small molecule ligands, such as drugs (Lin et al., 2022a; 2023a; Hu et al., 2022). Molecular docking which reveals this interaction, is critical in drug design and involves predicting the conformation of a protein-ligand complex. A key challenge in molecular docking lies in the *induced-fit* mechanism (Sherman et al., 2006), where the protein's binding sites (pockets) are *flexible*, altering their poses in response to ligand binding. Notably, the pocket's sidechain atoms exhibit the most significant flexibility (Clark et al., 2019).




(a) Flexible docking and *priori* leakage (b) Steric clashes

*Figure 1.* The illustration of our motivation. (a) The dotted line represents current protocols take docked pockets as input, which are not accessible in realistic scenarios and provide hints or leakage for ligand poses prediction. (b) The result of DiffDock (pdb_id: 6nsv); the steric clashes are highlighted with red circles, where the ligand overlaps with the protein surface (i.e., sidechains).

Though important, accurately predicting the bound structures is highly challenging. Traditional docking methods (Eberhardt et al., 2021; Alhossary et al., 2015), using empirical scoring functions and optimization algorithms, struggle with the vast search space and complex calculations, often resulting in inaccurate and slow predictions. Recent deep learning approaches (Zhou et al., 2023; Liao et al., 2019) focus on predicting ligand binding poses using bound protein structures (holo-structures), which include ground truth sidechain conformations (i.e. *priori* leakage). These approaches, by introducing ***priori* leakage** (as illustrated in Fig. 1(a)), oversimplify the problem and fail to mimic realistic docking scenarios. Other deep learning methods (Pei et al., 2023; Corso et al., 2023) ignore sidechains for simplicity and implicit flexibility. **This neglect often results in unrealistic poses** where ligands overlap with side chains, i.e. steric clashes (as shown in Fig.1(b)). More accurate and realistic binding pose predictions require explicit sidechain modeling. Additionally, protein-molecule interaction is the key to model the binding process. Current approaches learn this interaction implicitly with neural networks while **fail to incorporate explicit modeling of interaction** in 3D coordinates for direct and accurate guidance.

To address these limitations, we introduce an under-explored

[*]Equal contribution [1]Zhejiang University, Hangzhou [2]AI Lab, Research Center for Industries of the Future, Westlake University [3]University of Washington, Seattle. Correspondence to: Stan Z. Li <cairi@westlake.edu.cn>.

*Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

task, named *flexible docking* (Sahu et al., 2024) to predict poses of ligand and pocket sidechains simultaneously and introduce Re-Dock, a flexible and Realistic generative Docking framework with explicit modeling of pocket sidechain flexibility and integrated interaction prior to steer the generation process. ReDock mimics the induced-fit process (Sherman et al., 2006) for *realistic docking scenarios* and generates *physically realistic conformations* by extending diffusion bridges (Wu et al., 2022a) to non-Euclidean manifolds of *implicit* geometries: rotations, translations, and torsion with *explicit* interaction prior in Euclidean space.

In detail, we construct neural diffusion models to imitate the bridge processes for generating flexible and realistic poses of both ligands and pockets. Diffusion bridges are stochastic processes that guarantee to yield given observations at the fixed terminal time (Liu et al., 2022). Notably, we model sidechain distributions autoregressively for better generation quality concerning their sequential nature (Zhang et al., 2023b). Unlike previous diffusion bridge processes defined over Euclidean space with molecular coordinates (Wu et al., 2022a), we explore bridge processes in geometric space, challenging for its implicit nature of modeling data points. For constructing interaction-informed prior bridges over geometry, we enable *Energy-to-Geometry mapping* using the Newton-Euler equations inspired by rigid body mechanics.

We benchmark *flexible docking* in the pocket-aware setting and provide generalized results of Re-Dock with pocket predictions. In many drug discovery pipelines, pockets are identified early (Zhang et al., 2023a) and their sidechains contribute to the majority of flexibility (Clark et al., 2019) during docking. Thus Re-Dock focuses on these flexible pocket sidechains. We design a new benchmark that reflects realistic scenarios, including apo crystal docking and cross-dock using the PDBBind (Liu et al., 2017) and our curated datasets. Re-Dock shows comparative results in overall benchmark tasks, proving its effectiveness in predicting flexible docking structures. This demonstrates its potential for real-world applications. Our key contributions are:

- We introduce the under-explored task of *flexible docking* and design a rigorous benchmark with new datasets.
- We propose Re-Dock, a novel diffusion bridge model extended to non-Euclidean manifolds with *energy-to-geometry mapping* inspired by mechanics. It enables interaction-aware, 'induced', generative docking processes with co-modeling of binding energy and poses.
- Superior benchmark test results including cross-dock suggest our potential for real-world applications.

## 2. Related Works

**Molecular Docking.** This field predicts how proteins and ligands bind together. Traditional approaches, like AutoDock Vina (Eberhardt et al., 2021; Alhossary et al., 2015), SMINA (Koes et al., 2013), and GLIDE (Halgren et al., 2004), utilize energy-based functions. Recently, deep learning, especially Graph Neural Network (Wu et al., 2022d; 2023c; Zheng et al., 2023b) has brought two innovations: Regression-based methods such as Equibind (Stärk et al., 2022), Tankbind (Lu et al., 2022), and Fabind (Pei et al., 2023) for predicting the ligand's docking pose directly and generative docking introduced by DiffDock (Corso et al., 2023; Plainer et al., 2023) approaching docking as the generation of ligand geometries, like rotations. Most methods assume known holo-protein structures, which is often impractical. Traditional approaches and some structure prediction methods (Qiao et al., 2023; Krishna et al., 2023) can account for protein flexibility but require extensive computations. Our work lies in generative docking with pocket flexibility and explicit modeling of interaction, thus can be applied to more realistic tasks, including cross-dock. More related works can be found in the Appendix. A.

**Diffusion Bridge Process.** Diffusion-based generative models (Song et al., 2020; Ho et al., 2020) have become popular in AI generation by adding noise to data and then using a reverse process to generate outputs. Efforts to improve these models have led to diffusion bridges. Schrodinger bridges (De Bortoli et al., 2021; Shi et al., 2023) have been proposed for learning entropy-regularized optimal transports of generation paths and guarantee desirable outputs, but these models involve iterative proportional fittings and are computationally costly. Research works such as (Peluchetti, 2021) and prior bridge (Wu et al., 2022a) directly learn diffusion trajectories with specific ending data points and inject problem-dependent prior into these paths, avoiding the time-reversal technique of (Song et al., 2020). However, applying diffusion bridges to geometric domains (e.g., rotations and torsion) requires designing appropriate bridge processes, which is an area that remains unexplored.

## 3. Backgrounds

### 3.1. Diffusion Bridge for Non-geometric Domains

In this section, we provide an overview of the general diffusion bridge to introduce the necessary notations and concepts based on (Wu et al., 2022a; Liu et al., 2022).

**Definition.** Diffusion bridges are diffusion processes that are conditioned to initialize and terminate at two given states. We can learn a generative model given a dataset $\{x^{(k)}\}_{k=1}^{n}$ drawn from an unknown distribution $\Pi^*$ on $\mathbb{R}^d$ via building and imitating such a diffusion bridge with prior and data distribution on both sides. A diffusion bridge generative model on the time interval [0,1] is:

$$\mathbb{P}^\theta: \quad \mathrm{d}Z_t = s_t^\theta(Z_t)\mathrm{d}t + \sigma_t(Z_t)\mathrm{d}W_t, \forall t \in [0,1], Z_0 \sim \mu_0. \tag{1}$$

where $W_t$ is a standard Wiener process; $\sigma_t : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is

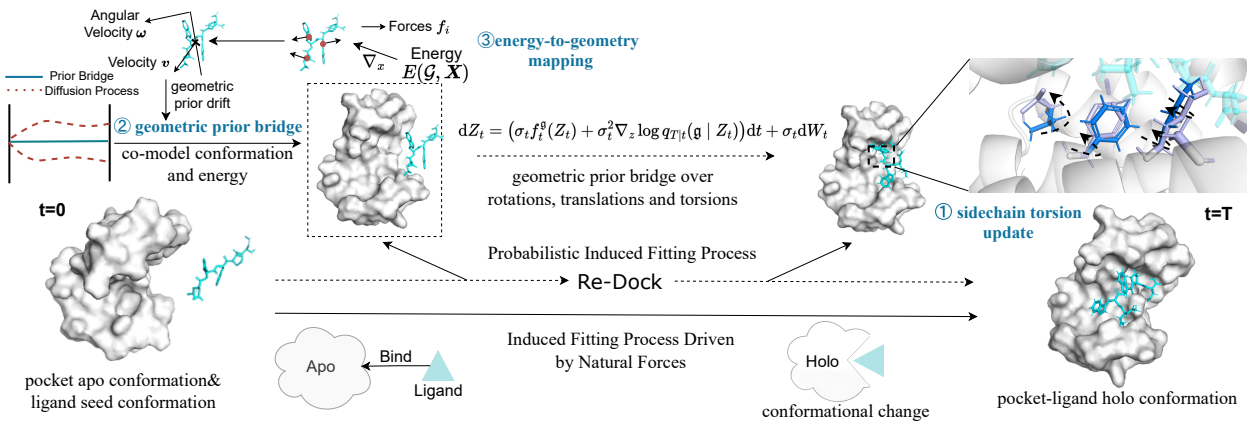



*Figure 2.* The illustration of `Re-Dock` Framework. We aim to simulate the induced fitting process with geometric prior bridges. Our key designs are threefold: ① The pocket sidechains displace the most flexibility for inducing interactions. Thus, we generate the sidechain conformations (the blue and purple sticks are the conformations of two steps before and after, respectively; we omit other sidechains for simplicity) via torsion angle updates while docking. ② We explore a novel generative model, the geometric prior bridge for reflecting the energy-constrained fitting process. Compared with the diffusion processes (the red curves), the prior bridge process (the blue line) is augmented with problem-dependent prior and thus more fast and accurate to generate. ③ For explicit modeling of interaction and constructing prior bridges over geometries, we propose an energy-to-geometry mapping module inspired by Newton-Euler equations.

a positive definition covariance coefficient; $s_t^\theta : \mathbb{R}^d \to \mathbb{R}^d$ is parameterized as a neural network with parameter $\theta$, and $\mu_0$ is the prior distribution. Here, $\mathbb{P}^\theta$ represents the distribution of the diffusion trajectory $Z = \{Z_t : t \in [0,1]\}$ and $\mathbb{P}_t^\theta$ is the marginal distribution of $Z_t$ at time $t$. We aim at learning a generative model with parameter $\theta$ such that the distribution $\mathbb{P}_1^\theta$ of the terminal state $Z_1$ equals the data distribution $\Pi^*$.

**Building and Learning Diffusion Bridges.** There are infinitive diffusion processes $\mathbb{P}^\theta$ that reach the same terminal distribution but differ in distributions of trajectories $Z$. Diffusion Bridge has the appealing potential to inject problem-dependent prior information into the trajectories and learning processes to obtain a generative model $\mathbb{P}^\theta$ fast and accurately. To achieve this, (Wu et al., 2022a) elicit an *imputation* process $\mathbb{Q}^x$ for each $x \in \mathbb{R}^d$, such that a draw $Z \in \mathbb{Q}^x$ yields trajectories that 1) are pinned at $Z_1 = x$, and 2) reflect important physical prior information on the current problem with more physically regularized processes.

Formally, if $\mathbb{Q}^x(Z_1 = x) = 1$, then $\mathbb{Q}^x$ or simply an $x$-bridge is a bridge process ending at data point $x$. We can construct physically informed diffusion bridges based on $\mathbb{Q}^x$ if we first sample a data point $x \in \Pi^*$ and then draw a bridge $Z \in \mathbb{Q}^x$ pinned at $x$. Thus the distribution of trajectories $Z$ is a mixture of $\mathbb{Q}^x$: $\mathbb{Q}^{\Pi^*} := \int \mathbb{Q}^x(\cdot)\Pi^*(\mathrm{d}x)$.

We can learn the diffusion bridge model $\mathbb{P}^\theta$ by imitating the trajectories drawn from $\mathbb{Q}^{\Pi^*}$ since the crucial property of $\mathbb{Q}^{\Pi^*}$ is its end distribution equals the data distribution, i.e., $\mathbb{Q}_1^{\Pi^*} = \Pi^*$. This can be formulated by maximum likelihood or equivalently minimizing the KL divergence: $\min_\theta \left\{\mathcal{L}(\theta) := KL(\mathbb{Q}^{\Pi^*} \,||\, \mathbb{P}^\theta)\right\}$. Furthermore, assume that the bridge $\mathbb{Q}^x$ is a diffusion model of form:

$$\mathbb{Q}^x: \quad \mathrm{d}Z_t = b_t(Z_t \mid x)\mathrm{d}t + \sigma_t(Z_t)\mathrm{d}W_t, Z_0 \sim \mu_0. \quad (2)$$

where $b_t(Z_t|x)$ is an $x$-dependent drift term that needs to be carefully designed to meet the bridge condition and incorporate prior information simultaneously. We adopt a practical and simple family of bridges proposed by (Wu et al., 2022a) in this paper by introducing modifications to Brownian Bridges (Liu et al., 2022):

$$\mathbb{Q}^x_{\mathrm{bb},f} : \mathrm{d}Z_t = \left(\sigma_t f_t(Z_t) + \sigma_t^2 \frac{x - Z_t}{\beta_1 - \beta_t}\right)\mathrm{d}t + \sigma_t \mathrm{d}W_t. \quad (3)$$

where $\beta_t = \int_0^t \sigma_s^2 \mathrm{d}s$, and $f_t(Z_t)$ is an extra drift term which reflects physical prior (e.g., the physical force) and $Z_0 \in \mathcal{N}(x, \beta_1)$ or $\mathcal{N}(0, \beta_1)$ when $\beta_1$ is large enough (Liu et al., 2022). After choosing appropriate $\mathbb{Q}^x_{bb,f}$, using Girsanov theorem, the loss function can be reformed into a form of denoising score matching loss of (Heng et al., 2022):

$$\mathcal{L}(\theta) = \mathbb{E}_{Z\sim\mathbb{Q}^{\Pi^*}}\left[\frac{1}{2}\int_0^1 \Big\|\sigma(Z_t)^{-1}(s_t^\theta(Z_t) - b_t(Z_t \mid Z_1))\Big\|_2^2 \mathrm{d}t\right] + \text{const}. \quad (4)$$

where the const term contains the log-likelihood for the initial distribution $\mu_0$, which is a const in our problem.

### 3.2. Problem Statement

Given unbounded protein and ligand structures as inputs, *flexible docking* aims to predict the binding pose (i.e. atom coordinates) of the pocket sidechain and ligand, imitating the process of induced-fit. For simplicity and efficiency, we model the *changes* the ligand and sidechains undergo during binding following (Corso et al., 2023), instead of

directly modeling coordinates. Thus, we focus on an $(m+6)$-dimensional manifold, where $m$ represents the number of rotatable bonds, and six extra dimensions account for roto-translations of ligands relative to the fixed protein backbone.

Our approach differs from previous methods as we construct diffusion bridges $\mathbb{P}^\theta$ on geometric manifolds (e.g. rotation $SO(3)$ and torsion $SO(2)^m$) and integrate interaction priors into our *generative bridge training*. Next, we detail how to learn the probabilistic induced-fit process $\mathbb{P}^\theta$.

## 4. Method

In this section, we formally describe the flexible and Realistic protein-ligand Docking (Re-Dock) framework. We aim to model protein-ligand interactions more accurately by considering the flexibility of pocket sidechains and incorporating a novel geometric diffusion bridge model to inject interaction prior. Our method draws inspiration from recent advances in Diffusion Bridge (Liu et al., 2022; Wu et al., 2022a), which integrates problem-dependent prior into generative paths. However, constructing suitable bridge processes on geometry with interaction prior is challenging. We address this by extending the diffusion bridge to geometric sub-manifolds and incorporating interaction prior with the Newton-Euler Equation. The overall framework is introduced in Section.4.1, with further elaboration on details of our method and its learning process with energy-conformation co-modeling in Section.4.2 and Section.4.3, respectively. A high-level schematic is provided in Fig. 2.

### 4.1. Diffusion Prior Bridge on Geometries

The geometric sub-manifold $\mathcal{M}_C$ is a product space $\mathbb{G}$ of the 3D translation group $\mathbb{T}(3)$, the 3D rotation group $SO(3)$ of rigid rotations of the ligand, and the 2D rotation groups $SO(2)^m$ of changes in m torsion angles $\theta$ of the pocket-ligand complex. As a starting point, We first build a simple Brownian Bridge on $\mathbb{G}$ via Doob's h-transforms (Liu et al., 2022). It states the Brownian bridge $\mathbb{Q}^x_{bb}$ can be shown to be the law of:

$$\mathbb{Q}^x_{bb} : \mathrm{d}Z_t = \sigma_t^2 \nabla_z \log q_{T|t}(x \mid Z_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t. \quad (5)$$

where $q_{T|t}(x \mid Z_t)$ is the density function of the transition probability $\mathbb{Q}_{T|t}(\cdot|z) = \mathcal{N}(z, \beta_T - \beta_t)$,where $\beta_t = \int_0^t \sigma^2 \mathrm{d}s$. Since $\mathbb{G}$ is a product manifold, the forward diffusion proceeds independently in each manifold, and the tangent space is a direct sum: $T_g\mathbb{Q} = T_{\mathbf{r}}\mathbb{T}_3 \oplus T_R SO(3) \oplus T_{\boldsymbol{\theta}} SO(2)^m \cong \mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \mathbb{R}^m$ where g = $(\mathbf{r}, R, \boldsymbol{\theta})$. Thus, the construction of the bridge on $\mathbb{G}$ is equivalent to building Brownian Bridges on each geometry independently. In all three groups, we can define the Brownian Bridge as Equation 5 where $\sigma = \sigma^{\mathbf{r}}, \sigma^R, \sigma^\theta$ for $\mathbb{T}(3), SO(3), SO(2)^m$ respectively, $\mathcal{N}(\cdot,\cdot)$ is the corresponding *normal distribution* on geometry and W is the corresponding Brownian motion. As mentioned in Section.3.1, we can add an extra drift term $f_t(z_t)$ that reflects the interaction prior and design a prior bridge $\mathbb{Q}^{\mathfrak{g}}_{bb,f}$ where $\mathfrak{g} = \mathbf{r}, R, \boldsymbol{\theta}$ as:

$$\mathrm{d}Z_t = \left(\sigma_t f_t^{\mathfrak{g}}(Z_t) + \sigma_t^2 \nabla_z \log q_{T|t}(\mathfrak{g} \mid Z_t)\right)\mathrm{d}t + \sigma_t \mathrm{d}W_t. \quad (6)$$

Since $\mathbb{T}(3) \cong \mathbb{R}^3$, it is trivial to build the bridge on translation in the same form of Equation 3 with variance $\sigma_t^r$. The $SO(2)^m$ group is diffeomorphic to the torus $\mathbb{T}^m$, on which the diffusion kernel is a *wrapped normal distribution* (Jing et al., 2022) with variance schedule $\sigma_t^\theta$. The transition probability $q_{T|t}(\theta \mid \theta_t)$ is:

$$q_{T|t}(\boldsymbol{\theta}|\boldsymbol{\theta}_t) \propto \sum_{\boldsymbol{d}\in\mathbb{Z}^m} \exp\left(-\frac{\|\boldsymbol{\theta} - \boldsymbol{\theta}' + 2\pi \boldsymbol{d}\|^2}{\beta_T - \beta_t}\right) \quad (7)$$

This can be sampled directly, and the score can be precomputed as a truncated infinite series. We consider the isotropic Gaussian distribution on SO(3) (Leach et al., 2022). The transition kernel on $SO(3)$ is $\mathcal{IG}_{SO(3)}(R, \beta_T - \beta_t)$ distribution, which can be sampled in the axis-angle parameterization by sampling a unit vector $\hat{\omega} \in \mathfrak{so}(3)$ uniformly and random angle $\omega \in [0, \pi]$ according to:

$$p(\omega) = \frac{1 - \cos\omega}{\pi} f(\omega), \sigma = \beta_T - \beta_t,$$
$$f(\omega) = \sum_{l=0}^{\infty} (2l+1)\exp(-l(l+1)\sigma^2/2)\frac{\sin((l+1/2)\sigma^2}{\sin(\omega/2)} \quad (8)$$

The score computation and sampling can be accomplished efficiently by precomputing the truncated infinite series and interpolating the CDF of $p(\omega)$, respectively.

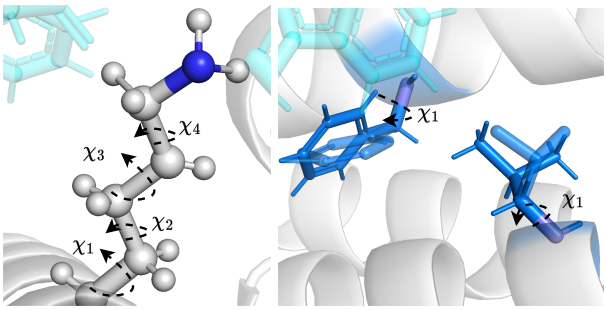
**Autoregressive Update on Sidechains.** Adding noises to (i.e. rotating) one angle $\boldsymbol{\theta}_i^{sc}$ will result in changes in and beyond the corresponding atom groups (for example, rotating $\theta_1^{sc}$ as in Fig. 3(b)). Thus, adding noise on all angles simultaneously will result in cumulative coordinate displacements in later angle atom groups and may damage the pocket structure, which complicate the denoising process of the latter angles (Zhang et al., 2023b).

Compared with torsion angles without a natural order in molecules, the sidechain angles have a natural sequential order (shown in Fig. 3(a)). To eliminate the cumulative effect and capture the pre- and post-dependencies when posing sidechains in complicated docking complex structures, we decompose the joint distribution of four sidechain torsional angles $\boldsymbol{\theta}^{sc}_{1,2,3,4}$ into individual conditional distributions:

$$p(\boldsymbol{\theta}^{sc}_{1,2,3,4}) = p(\boldsymbol{\theta}^{sc}_1) \cdot p(\boldsymbol{\theta}^{sc}_2|\boldsymbol{\theta}^{sc}_1) \cdot p(\boldsymbol{\theta}^{sc}_3|\boldsymbol{\theta}^{sc}_{1,2}) \cdot p(\boldsymbol{\theta}^{sc}_4|\boldsymbol{\theta}^{sc}_{1,2,3}). \quad (9)$$

This allows us to add noise to specific $\theta_i^{sc}$ angles in training iterations for simplifying denoising and generate the sidechain conformations step-by-step: we first predict $\boldsymbol{\theta}_1^{sc}$ for all residues based on the backbone and ligand structure. Then, using the complex with predicted $\boldsymbol{\theta}_1^{sc}$, we proceed to predict $\boldsymbol{\theta}_2^{sc}$, and so on for $\boldsymbol{\theta}_3^{sc}$ and $\boldsymbol{\theta}_4^{sc}$.

**Overall Framework.** We train generative bridge models $\mathbb{P}^\theta_{\mathfrak{g}}$ on each geometry by minimizing $\mathcal{L}_\theta$ in Eq.4 and obtain final prior bridge model $\mathbb{P}_\theta$ combining each bridge for probabilistic induced fitting.




(a) Sidechain Angles (b) Effect of rotating $\theta_1^{sc}$

*Figure 3.* The illustration of sidechain updates. (a) Up to four sidechain $\theta^{sc}$ angles (formally, $\chi$ angles) have a sequential order. (b) Rotating $\theta_1^{sc}$ will affect the coordinates of atoms in $\theta_2^{sc}$, $\theta_3^{sc}$ and $\theta_4^{sc}$. It's similar for rotating $\theta_2^{sc}$ and $\theta_3^{sc}$. The atom groups of the later angles will accumulate noise from the former angles which complicates the latter's denoising process.

To parameterize the bridge $\mathbb{P}^\theta$, we use a relation-aware equivariant graph neural network on full atoms of complexes as a shared encoder and the equivariant pooling layer, invariant pooling layer for predicting SE(3)-equivariant (i.e., translations and rotations) and invariant (i.e., binding energy, torsion angles) quantities. Algorithmic description and more details on model architecture, data modeling, and training can be referred in Appendix. B.

### 4.2. Injecting Interaction Prior into Geometry Bridges.

Although we have defined the prior bridges and score matching loss on $\mathbb{G}$, we nevertheless develop the model and integrated interaction prior in 3D coordinates $\mathbf{X}$ directly. Providing the full 3D structure, rather than abstract elements of the product space, to the score model allows it to reason about physical interactions using *SE(3)* equivariant model, and not be dependent on arbitrary definitions of torsion angles (Jin et al., 2023). Further, interaction priors defined on abstract elements of geometries are less informative, and incorporating physical constraints defined on 3D coordinates can provide direct and accurate guidance over interactions within pocket-ligand complexes. Such design can also allow seamless integration of well-defined physical or statistical potential energy priors of protein-ligand interactions, which we choose in this paper.

Bridging physical energy priors defined in 3D coordinates and prior bridges on geometry product space $\mathbb{G}$ is challenging as the prior drift terms in bridges are not directly comparable with the physical energy priors $E(\mathbf{X})$, nor the derivatives, i.e., physical forces $\mathbf{F} = \partial E/\partial \mathbf{X}$. To fill the gap and enjoy the benefits of both modeling schemes, we propose an *Energy-to-Geometry mapping* method inspired by rigid body mechanics with Newton-Euler equations.

**Energy-to-Geometry Mapping.** There are inherent connections between energy and rigid body motions in rigid body mechanics. Given a potential energy $E(\mathbf{X})$, we can calculate the forces $\mathbf{F} = \partial E/\partial \mathbf{X}$ acting on a rigid body system. Then, the system's combined translational and rotational dynamics can be described in the Newton-Euler equations (Eq. 11). It is natural to take inspiration from rigid body mechanics to convert prior energy defined in 3D coordinates to corresponding geometric updates as our extra drift terms $f_t^{\mathfrak{g}}(Z_t)$ in geometric prior bridges.

Too short distances or clashes between atoms of ligands and pockets in the complex conformations generation process will induce abnormal Van der Waals forces. Thus, we incorporate anti-clash potential prior in our bridges. Following (De Bortoli et al., 2021; Corso et al., 2023), we choose $\{\boldsymbol{x} \in \mathbb{R}^3 : S(\boldsymbol{x}) = \gamma\}$ where $S(\boldsymbol{x}) = -\sigma \ln\left(\sum_{j=1}^{N_P} \exp\left(-\|\boldsymbol{x} - \boldsymbol{x}_P^j\|^2/\sigma\right)\right)$ as the descriptor of the protein surface. $N_P$ is the atom number of pockets and $\{\boldsymbol{x}_P^j\}_{j=1}^{N_P}$ is the set of pocket atom coordinate vectors. The anti-clash potential we used can be derived as follows:

$$E_{clash}(\boldsymbol{X}) = -\sum_{i}^{N_M} \max\left(0, \gamma - S(\boldsymbol{x}_t^{(i)})\right), \tag{10}$$

where $N_M$ is the atom number of molecules. We also include the *Amber* physical interaction energy terms as prior potentials. The overall interaction potential energy prior $E(\boldsymbol{X})$ is a direct sum detailed in the Appendix. D.

**Newton-Euler Equation** The Newton-Euler equations explain a rigid body's combined translational and rotational dynamics. In the Center of Mass (CoM) frame, this can be written in matrix form as:

$$\begin{pmatrix} \boldsymbol{F} \\ \boldsymbol{\tau} \end{pmatrix} = \begin{pmatrix} m\mathbf{I} & 0 \\ 0 & \mathbf{I}_c \end{pmatrix} \begin{pmatrix} d\boldsymbol{v}/dt \\ d\boldsymbol{\omega}/dt \end{pmatrix} + \begin{pmatrix} 0 \\ \boldsymbol{\omega} \times \mathbf{I}_c \boldsymbol{\omega} \end{pmatrix} \tag{11}$$

where $\boldsymbol{F}$ and $\boldsymbol{\tau}$ are the total force and torque acting on CoM, $\boldsymbol{v}$ and $\boldsymbol{\omega}$ are the velocity of CoM and the angular velocity around CoM, $m$ and $\mathbf{I}_c$ is the mass and inertia matrix of the rigid body, which are constant for a given rigid body.

In our *Energy-to-Geometry mapping* module, the force $\boldsymbol{f}_t^i$ act on each ligand atom $i$ is defined as the gradient of the interaction energy prior function $\boldsymbol{f}_t^i = \partial E(\boldsymbol{X_t})/\partial \boldsymbol{x}_t^i$. We can then calculate the total force $\boldsymbol{F_t}$ and torque $\boldsymbol{\tau}_t$ for the ligand in the discrete-time step $t$:

$$\boldsymbol{F}_t = \sum_{i \in \mathcal{V}_L} \boldsymbol{f}_t^i, \boldsymbol{\tau}_t = \sum_{i \in \mathcal{V}_L} (\boldsymbol{x}_t^i - \boldsymbol{p}_t^c) \times \boldsymbol{f}_t^i. \tag{12}$$

where $\boldsymbol{p}_t^c$ is the Center of Mass. We assume the system is stationary at each discrete time step $t$ (Guan et al., 2023; Jin et al., 2023). Furthermore, we can specify changes of torsion angles to be *disentangled* from rotations or translations (Corso et al., 2023) to ensure torsional updates cause no linear or angular momentum, i.e., $\boldsymbol{\omega}_t = 0$ and $\boldsymbol{v}_t = 0$. Thus, the Newton-Euler equations (Eq.11) can be simplified as $\boldsymbol{F}_t = m\frac{d\boldsymbol{v}_t}{dt}$ and $\boldsymbol{\tau}_t = \boldsymbol{I}_c \frac{d\boldsymbol{\omega}_t}{dt}$. For a short enough time period $\Delta t$, we have the velocity and angular velocity of

| | Holo Crystal Proteins | | | | | | Apo ESMFold Proteins | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Top-1 RMSD | | | Top-5 RMSD | | | Top-1 RMSD | | | Top-5 RMSD | | | Average |
| Method | %<2 | %<5 | Med. | %<2 | %<5 | Med. | %<2 | %<5 | Med. | %<2 | %<5 | Med. | Runtime (s) |
| GNINA | 34.6 | 62.6 | 3.3 | 46.8 | 78.1 | 2.1 | 9.9 | 31.7 | 6.7 | 18.1 | 53.5 | 4.6 | 251 |
| SMINA | 21.5 | 50.1 | 5.0 | 37.2 | 70.8 | 2.8 | 4.9 | 19.2 | 7.3 | 9.6 | 43.4 | 5.4 | 246 |
| LeDock | 24.0 | 48.6 | 5.1 | 39.4 | 69.8 | 2.9 | 10.9 | 36.7 | 6.2 | 19.5 | 59.0 | 4.2 | 7 |
| SurfLex | 6.1 | 31.9 | 6.7 | 10.6 | 53.1 | 4.7 | 1.4 | 15.6 | 7.4 | 2.6 | 30.1 | 6.1 | 5 |
| QVina | 21.3 | 50.9 | 4.9 | 36.8 | 71.8 | 2.9 | 5.0 | 19.5 | 7.3 | 9.9 | 42.9 | 5.4 | 2 |
| GNINA-Flex | 29.4 | 63.7 | 3.5 | 45.7 | 84.9 | 2.2 | 10.9 | 36.7 | 6.2 | 19.5 | 59.0 | 4.2 | 1072 |
| SMINA-Flex | 20.5 | 46.2 | 5.3 | 34.4 | 73.3 | 2.9 | 5.8 | 18.4 | 7.1 | 11.1 | 42.0 | 5.4 | 963 |
| FlexPose | 42.0 | 79.3 | 2.2 | - | - | - | 34.8 | **79.9** | 2.8 | - | - | - | 12 |
| DiffDock(pocket, 40) | 51.8 | 75.3 | 2.0 | 60.7 | 79.2 | 1.9 | 37.8 | 72.4 | 2.6 | 38.1 | 73.4 | 2.5 | 61 |
| **Re-Dock (10)** | 51.9 | 77.4 | 2.0 | 62.2 | 83.6 | 1.7 | 39.0 | 74.8 | 2.5 | 39.0 | 74.8 | 2.5 | 15 |
| **Re-Dock (40)** | **53.9** | **80.3** | **1.8** | **65.0** | **86.7** | **1.4** | **42.9** | 76.4 | **2.4** | **45.6** | **78.2** | **2.2** | 58 |

*Table 1.* Performance of flexible redocking on the PBDBind test set (bound, holo-crystal proteins) and its corresponding ESMFold predicted Apo structures (unbound). The best metrics are marked by **bold**. In parenthesis, we specify the number of poses sampled from the generative model. It is worth noting that only our `Re-Dock` has no access to the holo-crystal proteins. Our `Re-Dock` surpasses all baselines across different metrics and settings with affordable inference time, which demonstrates our effectiveness and efficiency on flexible redocking as well as apo docking, and the advantages of our geometric prior bridge with sidechain flexibility.

the ligand as $\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{I}_c^{-1}\boldsymbol{\tau}_t\Delta t$ and $\boldsymbol{v}_{t+\Delta t} = \frac{1}{m}\boldsymbol{F}_t\Delta t$. Assuming each atom in the ligand has the unit mass and setting the time period $\Delta t$ as hyper-parameter $\alpha$, the prior drift terms in prior geometric bridges $\mathbb{Q}^{\mathfrak{g}}_{bb,f}$(Eq. 6) can be designed as follows:

$$f_t^{\tau}(Z_t) = \frac{\alpha}{|\mathcal{V}_L|}\boldsymbol{F}_t, f_t^R(Z_t) = \alpha\boldsymbol{I}_c^{-1}\boldsymbol{\tau}_t \tag{13}$$

where $|\mathcal{V}_L|$ denotes the number of atoms in the ligand. More details and analysis (e.g., the equivariance analysis and $f_t^{\boldsymbol{\theta}}$) can be referred to the Appendix. D.

### 4.3. Learning by Co-Modeling Energy and Poses

We have designed interaction-aware prior bridges on geometries via Eq.6 and Eq.13. Next we will discuss how to better parameterize the learnable drift $s_t^\theta$ in Eq.1 for learning the bridges. Specifically, we assume the learnable drift has a form of $s_t^\theta = \beta\hat{f}_t + \hat{s}_t^\theta$ where $\beta$ can be another learnable parameter, $\hat{s}_t^\theta$ is direct predicted scores and $\hat{f}_t$ is also calculated using *Energy-to-Geometry mapping* with *learnable* interaction energy $E^\theta(\boldsymbol{X}, \mathcal{G})$ as in equation 13.

This parameterization allows learning of prior drifts and the binding energy via an additional head at the same time. After training, $E^\theta(\boldsymbol{X}, \mathcal{G})$ can also serve as the *confidence model* (Corso et al., 2023) in addition to calculating $s_t^\theta$.

This idea has a connection with the Energy-Based model (EBMs) (Jin et al., 2023), where the likelihood of a data point $p(\boldsymbol{X}, \mathcal{G}) \propto \exp(-E^\theta(\boldsymbol{X}, \mathcal{G}))$. Our training process can be seen as training EBMs using denoising score matching with prior, and then we can interpret the learned energy as protein-ligand interaction energy, naturally including the interaction potential energy prior.

Our method combines parameterizing the score directly (leaving the energy function *implicit*) and *explicitly* modeling the energy function. Thus, we enjoy both advantages of sampling quality and likelihood estimation to co-model energy and poses for imitating energy-constrained **induced-fit** process using denoising score matching loss in Eq. 4.

## 5. Experiments

In this section, we justify the advantages of the proposed `Re-Dock` with comprehensive experiments. The experimental setup is introduced in Section 5.1. We aim to answer six research questions. **Q1:** How effective is `Re-Dock` for flexible re-docking and apo-docking tasks? **Q2:** Can `Re-Dock` generate accurate sidechain conformations? **Q3:** Can `Re-Dock` generalize well to more challenging and realistic tasks (e.g. cross-dock)? **Q4:** How do key framework designs impact the performance of `Re-Dock`? Can `Re-Dock` use fewer sampling steps to further speed up large-scale screening? **Q5:** Can `Re-Dock` generate physically valid samples? **Q6:** What's the quality of generated samples of `Re-Dock`?

### 5.1. Experimental Setups

To construct a rigorous and reasonable benchmark for flexible docking, we design and set up four tasks with increasing modeling flexibility and difficulty. As mentioned above, we focus on results in the pocket-aware setting. Pockets are defined as residues within 8 Å of ligands. Details and generalized results with pocket predictions are in the Appendix.

**Evaluation Tasks.** Tasks include: a) **Flexible re-docking**, modified from standard re-docking tasks (Corso et al., 2023; Pei et al., 2023). Instead of using holo protein structures as input, we randomize pocket sidechain conformations at first and require the docking models to predict bound ligand poses based on noisy pockets. b) **Apo-dock**, which takes the unbound pocket structures as input. This task aligns with realistic scenarios, and we conduct it on the *crystal apo protein structures* in addition to their *computational alternatives* used in previous works (Corso et al., 2023). c) **Sidechain pose prediction**. We also need to generate pocket

| | Holo Crystal Proteins | | | | | Apo ESMFold Proteins | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Top-1 SC-RMSD | | | Top-5 SC-RMSD | | Top-1 SC-RMSD | | | Top-5 SC-RMSD | |
| Method | %<1 | %<2 | Med. | %<1 | Med. | %<1 | %<2 | Med. | %<1 | Med. |
| GNINA-Flex | 3.3 | 71.9 | 1.7 | 7.7 | 1.4 | 0.6 | 31.0 | 2.5 | 1.8 | 2.0 |
| SMINA-Flex | 2.0 | 63.8 | 1.8 | 8.3 | 1.4 | 0.6 | 34.4 | 2.4 | 1.8 | 2.0 |
| **Re-Dock (10)** | 86.8 | 95.4 | 0.6 | 93.1 | 0.6 | **39.8** | 80.2 | 1.2 | 45.7 | 1.1 |
| **Re-Dock (40)** | **87.5** | **100.0** | **0.6** | **93.6** | **0.6** | 38.4 | **82.5** | **1.2** | **46.5** | **1.1** |

*Table 2.* Performance of sidechain generation on PBDBind test set and its corresponding ESMFold predicted Apo structures. The best metrics are marked by **bold**. In parenthesis, we specify the number of poses sampled from the generative model. The results suggest our superior sidechain generation performance.

| | Top-1 RMSD | |
|---|---|---|
| Methods | %<2 | Med. |
| DiffDock (pocket, 20 sampling steps, 40) | 51.8 | 2.0 |
| Re-Dock (20 sampling steps, 40) | **53.9** | **1.8** |
| w/o Sidechain Generation | 39.8 | 3.3 |
| w/o Prior Energy | 50.3 | 2.4 |
| DiffDock (pocket, interaction prior guided sampling, 40) | 52.7 | 1.9 |
| Re-Dock (10 sampling steps, 40) | 52.1 | 2.0 |

*Table 3.* Ablation study for designed components on flexible redocking task in PDBBind test set. The best metrics are marked by **bold**. In parenthesis, we specify the number of poses sampled from the generative model. w/o Sidechain Generation refers to removing the sidechain torsion sampling process in inference. w/o Prior Energy means training a vanilla geometric bridge without prior drift. Interaction prior guided sampling adds classifier guidance to DiffDock(pocket) using *Energy-to-Geometry mapping*.

sidechain poses to better understand interactions. Similar to the ligand pose prediction, we formulate this task to assess generated samples. d) **Cross-dock**. It's an important real-world task for drug discovery (Zhang et al., 2023a). The task involves predicting the ligand poses using pocket structures that are bound with different ligands, which can be biased and misleading.

**Datasets.** We conduct training, flexible redocking, and sidechain pose prediction evaluation on the PDBBind v2020 dataset (Liu et al., 2017) with the time-based dataset split following previous works (Corso et al., 2023; Pei et al., 2023). For apo-docking, we use the ESMFold on the test set of PDBBind to obtain predicted apo structures. We also curate corresponding apo crystal structures with the PDBBind test via searching against Protein Data Bank (Burley et al., 2021). For cross-dock, we collect complex structures bound with different ligands for 60 important drug targets and select 7 representative targets for evaluation, resulting in over 10000 complexes.

**Metrics.** We use Ligand RMSD as the evaluation metric, which calculates the root-mean-square deviation between the predicted and the holo crystal ligand atomic Cartesian coordinates. Following (Corso et al., 2023), we report the percentage of predictions with *RMSD* below 2 and 5, the median `RMSD` (Med.) and average runtime per complex. For sidechain pose prediction, we use a similar metric named SC-RMSD and report results with different thresholds of 1 and 2. For generated structures, the perfect match with the crystal structures is unrealistic, as they differ in bond lengths. To compensate for this fact, we use a relative measure that compares the SC-RMSD before and after the prediction.

**Baselines.** We compare `Re-Dock` with state-of-the-art search-based methods SMINA (Koes et al., 2013), GNINA (McNutt et al., 2021), LeDock (Zhao & Caflisch, 2013), SurfLex (Jain, 2003), and Qvina (Alhossary et al., 2015), and the recent deep learning method DiffDock (Corso et al., 2023). We implemented and retrained DiffDock(pocket) with pockets as input for comparison. We also include flexible docking baselines GNINA-Flex, SMINA-Flex, and FlexPose (Dong et al., 2023). For generative models, we sample various poses per complex and take the most accurate pose out of 1 or 5 highest-ranked predictions (Top-1 or Top-5) according to the confidence model.

Extensive details about the experimental setup including data curation and training can be found in the Appendix. C.

### 5.2. Flexible Redocking and Apo-docking (Q1)

In the context of flexible redocking, we observe the notable performance of our `Re-Dock`. Across all metrics, `Re-Dock` ranks as the best as presented in the left half of Table 1. It is worth noting that, all baselines in the flexible redocking task take the holo pocket structure as input for they are not trained for or perform much worse on noisy structures, while our `Re-Dock` receive randomized pocket sidechain poses which are more challenging. Detailed results can be referred in the Appendix. C. For Apo docking on both ESMFold (Lin et al., 2022b) predicted structures and apo crystal structures, our `Re-Dock` also has leading performances as shown in the right half of Table 1 and Table 4. The results suggest the advantage of our geometric prior bridge and explicit sidechain pose modeling.

### 5.3. Quality of Sidechain Pose Generation (Q2)

Table 2 shows the results of sidechain pose prediction in the PDBBind test set and corresponding apo structures. Our `Re-Dock` significantly outperforms all flexible docking baselines, especially in the metric of the percentage of SC-RMSD $< 1$, highlighting the fine-grain modeling ability of our method. Different from the flexible redocking task above, all methods in the sidechain pose prediction task

| Method | CrossDock Top-5 RMSD %<2 | %<5 | Med. | Apo Crystal Proteins Top-1 RMSD %<2 | %<5 | Med. | Top-5 RMSD %<2 | %<5 | Med. |
|---|---|---|---|---|---|---|---|---|---|
| GNINA | 23.5 | 75.0 | 3.4 | 7.0 | 20.7 | 8.6 | 8.9 | 29.6 | 6.9 |
| SMINA | 16.8 | 72.2 | 3.7 | 5.6 | 15.3 | 8.7 | 8.4 | 27.3 | 7.2 |
| LeDock | 10.3 | 65.3 | 4.3 | 3.3 | 12.6 | 8.7 | 5.9 | 27.5 | 6.7 |
| SurfLex | 7.8 | 61.0 | 4.4 | 1.1 | 13.3 | 8.0 | 1.8 | 18.1 | 7.3 |
| QVina | 17.3 | 72.2 | 3.4 | 6.4 | 16.0 | 8.6 | 9.2 | 28.4 | 7.3 |
| GNINA-Flex | 19.6 | 81.1 | 3.0 | 7.0 | 19.6 | 8.3 | 9.8 | 35.7 | 6.5 |
| SMINA-Flex | 26.2 | 78.2 | 3.1 | 5.4 | 18.5 | 8.2 | 9.2 | 33.1 | 6.4 |
| $FlexPose^*$ | 13.8 | 77.4 | 3.7 | 23.3 | **61.7** | 3.9 | - | - | - |
| DiffDock(pocket) | - | - | - | 30.3 | 50.6 | 3.9 | 33.7 | 51.2 | 3.7 |
| **Re-Dock (10)** | **36.1** | **91.8** | **2.6** | 31.3 | 51.7 | 3.9 | 35.6 | 54.9 | 3.6 |
| **Re-Dock (40)** | - | - | - | **34.4** | 52.5 | **3.7** | **38.6** | **59.2** | **3.3** |

*Table 4.* Performance comparison on apo crystal re-docking and a more challenging task, cross-dock. The best metrics are marked by **bold** and we only generate 10 samples per complex for cross-dock to reduce computational burden. Methods with * only have Top-1 results. In addition to predicted approximated apo-structures, we provide promising results on apo crystal structures which are frequently used in real-world applications. Cross-dock requires the docking methods to predict correct ligand poses based on pocket structure bound with another ligand, which can be misleading and thus challenging. The results demonstrate the effectiveness and advantages of explicit modeling of flexible sidechains and our geometric prior bridge.

receive randomized pocket sidechain poses.

### 5.4. Performace Comparison on Cross-dock (Q3)

We test our `Re-Dock` on the challenging cross-docking task as shown in Table 4. Our method has achieved state-of-the-art performance on the benchmark test set, suggesting the comparative ability of `Re-Dock` on flexible docking. Our scheme of co-modeling pocket sidechain and ligand flexibility in a probabilistic induced fitting process contributes to the promising generalization ability over realistic docking scenarios and real-world applications. Since the size of the dockings per pocket is unevenly distributed, we report the pocket-level cross-docking performance via a pocket-normalized score (Brocidiacono et al., 2023).

### 5.5. Ablation Study (Q4)

As table 3 suggests, components designed for `Re-Dock` all contribute to our superior performance. The results highlight the contribution of sidechain generation and prior energy. Notably, modifying the training process with a prior bridge yields better results than using classifier guidance (Dhariwal & Nichol, 2021) in sampling. Further, we achieve better results than diffusion with fewer time steps.

### 5.6. More Results on the PoseBuster Benchmark (Q5)

Table 5 reports more results on PoseBuster (Buttenschoen et al., 2024) benchmark. We achieve the best performance over deep learning-based methods especially in the PB-Valid check setting, demonstrating the superior ability of geometric prior bridge for generating realistic samples.

### 5.7. Case Study (Q6)

**`Re-Dock` can generate 'induced' sidechain poses in realistic scenarios.** The first and last column of Fig.4 shows

| Methods | Top-1 RMSD %<2 | %<2&PB-Valid |
|---|---|---|
| Gold (Verdonk et al., 2003) | 58.0 | 55.0 |
| Vina (Eberhardt et al., 2021) | 60.0 | 58.0 |
| DeepDock (Liao et al., 2019) | 20.0 | 5.2 |
| Uni-Mol (Zhou et al., 2023) | 22.0 | 2.0 |
| TankBind (Lu et al., 2022) | 16.0 | 3.3 |
| DiffDock(40) | 38.0 | 12.0 |
| **Re-Dock**(40) | **50.7** | **32.8** |

*Table 5.* Flexible redocking results on PoseBuster benchmark. PB-Valid means the generated docking samples pass the PoseBuster Valid check and are physically valid (Buttenschoen et al., 2024). The best results among deep learning-based methods are **bold**. Results demonstrate the effectiveness of `Re-Dock` for generating realistic (i.e., physically valid) samples.

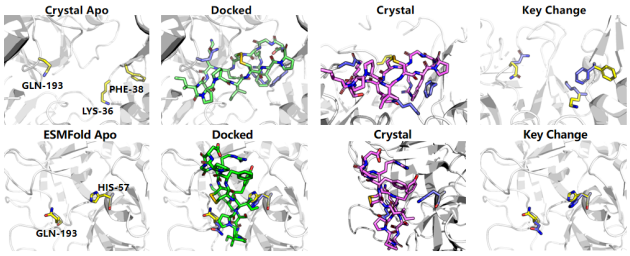



*Figure 4.* Case Studies on Plasmin (Gallus & Hirsh, 1976), a real-world protein target and its inhibitor ligand SFTI-1 (de Veer et al., 2021). The numbers (e.g. HIS-57) indicate the name and id of residues. `Re-Dock` replicates the bound crystal conformations for ligands (Docked) and sidechains (Key Change) on the apo (Top, PDB: 1QRZ) and predicted (Bottom) structure of plasmin.

`Re-Dock` can prompt the inward folding of residues that facilitate the accurate positioning of the SFTI-1 ligand in realistic scenarios using crystal apo or predicted structures.

**`Re-Dock` generates valid ligand pose for real-world targets** In the two middle columns of Fig.4, the real-world target Plasmin (PDB: 1QRZ) is not seen during training, but we can successfully generalize to this real-world challenge, predicting the accurate binding poses.

## 6. Conclusion

We introduce an under-explored docking task, flexible docking, which better aligns with realistic applications. Tailored to this task, we present `Re-Dock`, a geometric prior bridge generative model with the energy-to-geometry mapping inspired by the Newton-Euler equation in mechanics. We extend the prior bridge to the geometric manifold and propose a novel and general module that bridges the explicit interaction energy with implicit geometries to design geometric prior drifts and co-model the energy and poses. With explicit generative modeling of ligands and pocket sidechain poses, `Re-Dock` simulates the induced fitting process with a probabilistic one. The results on the proposed thoughtful benchmarks demonstrate our efficiency and effectiveness. Limitations still exist, including insufficient exploration of pocket prediction and end-to-end blind docking.

## Acknowledgements

This work was supported by the Science & Technology Innovation 2030 Major Program Project No. 2021ZD0150100, National Natural Science Foundation of China Project No. U21A20427, Project No. WU2022A009 from the Center of Synthetic Biology and Integrated Bioengineering of Westlake University, and Project No. WU2023C019 from the Westlake University Industries of the Future Research. Besides, we thank the help of Dr. Tailin Wu, and his great efforts in his deep insights into cutting-edge issues, the guidance provided in the rebuttal, and the funding that supplied us with the necessary equipment for our research. Finally, we thank the Westlake University HPC Center for providing part of the computational resources.

## Impact Statement

Unlike previous methods, `Re-Dock` is designed and trained for flexible docking and thus better generalizes to realistic applications with real-world drug targets and realistic results. It integrates binding energy prediction and can offer great value for many realistic drug discovery and protein engineering pipelines as well as disease understanding. With a better trade-off of efficiency and effectiveness (for instance, choosing a proper number of samples generated per complex and sampling steps would help), `Re-Dock` can be applied for both large-scale virtual screening and deeper analysis of protein-ligand interaction. While there exists the potential risk that docking methods could be misused to develop harmful drugs, it's important to note that drug development is subject to stringent oversight globally. This rigorous regulatory environment ensures that such misuses can be effectively managed and controlled.

## References

Aggarwal, R., Gupta, A., and Priyakumar, U. Apobind: a dataset of ligand unbound protein conformations for machine learning applications in de novo drug design. *arXiv preprint arXiv:2108.09926*, 2021.

Alhossary, A., Handoko, S. D., Mu, Y., and Kwoh, C.-K. Fast, accurate, and reliable molecular docking with quickvina 2. *Bioinformatics*, 31(13):2214–2216, 2015.

Brocidiacono, M., Popov, K. I., Koes, D. R., and Tropsha, A. Plantain: Diffusion-inspired pose score minimization for fast and accurate molecular docking. *ArXiv*, 2023.

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., et al. Rcsb protein data bank: powerful new tools for exploring 3d structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic acids research*, 49(D1):D437–D451, 2021.

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*, 2024.

Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., et al. *Amber 2021*. University of California, San Francisco, 2021.

Chen, R., Li, L., and Weng, Z. Zdock: an initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics*, 52(1):80–87, 2003.

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. *International Conference on Learning Representations (ICLR)*, 2023.

De Bortoli, V., Thornton, J., Heng, J., and Doucet, A. Diffusion schrödinger bridge with applications to score-based generative modeling. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/

940392f5f32a7ade1cc201767cf83e31-Paper.pdf.

de Veer, S. J., White, A. M., and Craik, D. J. Sunflower trypsin inhibitor-1 (sfti-1): sowing seeds in the fields of chemistry and biology. *Angewandte Chemie International Edition*, 60(15):8050–8071, 2021.

De Vries, S. J., Van Dijk, M., and Bonvin, A. M. The haddock web server for data-driven biomolecular docking. *Nature protocols*, 5(5):883–897, 2010.

DeLano, W. L. et al. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr*, 40(1): 82–92, 2002.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Dong, T., Yang, Z., Zhou, J., and Chen, C. Y.-C. Equivariant flexible modeling of the protein–ligand binding pose with geometric deep learning. *Journal of Chemical Theory and Computation*, 19(22):8446–8459, 2023.

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. doi: 10.1021/acs.jcim.1c00203. URL https://doi.org/10.1021/acs.jcim.1c00203. PMID: 34278794.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pp. 2021–10, 2021.

Gallus, A. S. and Hirsh, J. Prevention of venous thromboembolism. In *Seminars in thrombosis and hemostasis*, volume 2, pp. 232–290. Copyright© 1976 by Thieme Medical Publishers, Inc., 1976.

Ganea, O.-E., Huang, X., Bunne, C., Bian, Y., Barzilay, R., Jaakkola, T., and Krause, A. Independent se (3)-equivariant models for end-to-end rigid protein docking. *arXiv preprint arXiv:2111.07786*, 2021.

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

Guan, J., Peng, X., Jiang, P., Luo, Y., Peng, J., and Ma, J. Linkernet: Fragment poses and linker co-design with 3d equivariant diffusion. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=6EaLIw3W7c.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759, 2004.

Hartmann, C., Antes, I., and Lengauer, T. Docking and scoring with alternative side-chain conformations. *Proteins: Structure, Function, and Bioinformatics*, 74(3):712–726, 2009.

Heng, J., Bortoli, V. D., Doucet, A., and Thornton, J. Simulating diffusion bridges with score matching, 2022.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Hu, B., Xia, J., Zheng, J., Tan, C., Huang, Y., Xu, Y., and Li, S. Z. Protein language models and structure prediction: Connection and progression, 2022.

Hu, B., Tan, C., Xia, J., Zheng, J., Huang, Y., Wu, L., Liu, Y., Xu, Y., and Li, S. Z. Learning complete protein representation by deep coupling of sequence and structure. *bioRxiv*, 2023. doi: 10.1101/2023.07.05.547769. URL https://www.biorxiv.org/content/early/2023/07/07/2023.07.05.547769.

Huang, Y., Li, S., Su, J., Wu, L., Zhang, O., Lin, H., Qi, J., Liu, Z., Gao, Z., Liu, Y., et al. Protein 3d graph structure learning for robust structure-based protein property prediction. *arXiv preprint arXiv:2310.11466*, 2023a.

Huang, Y., Wu, L., Lin, H., Zheng, J., Wang, G., and Li, S. Z. Data-efficient protein 3d geometric pretraining via refinement of diffused protein structure decoy, 2023b.

Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, 46(4):499–511, 2003.

Jin, W., Sarkizova, S., Chen, X., Hacohen, N., and Uhler, C. Unsupervised protein-ligand binding energy prediction via neural euler's rotation equation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=46gYakmj4e.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *arXiv preprint arXiv:2206.01729*, 2022.

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.

Kong, X., Huang, W., and Liu, Y. Conditional antibody design as 3d equivariant graph translation. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=LFHFQbjxIiP.

Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G. R., Morey-Burrows, F. S., Anishchenko, I., Humphreys, I. R., et al. Generalized biomolecular modeling and design with rosettafold all-atom. *bioRxiv*, pp. 2023–10, 2023.

Krivák, R. and Hoksza, D. P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of cheminformatics*, 10:1–12, 2018.

Leach, A., Schmon, S. M., Degiacomi, M. T., and Willcocks, C. G. Denoising diffusion probabilistic models on so (3) for rotational alignment. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022.

Liao, Z., You, R., Huang, X., Yao, X., Huang, T., and Zhu, S. Deepdock: enhancing ligand-protein interaction prediction by a combination of ligand and structure information. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 311–317. IEEE, 2019.

Lin, H., Huang, Y., Liu, M., Li, X., Ji, S., and Li, S. Z. Diffbp: Generative diffusion of 3d molecules for target protein binding. *arXiv preprint arXiv:2211.11214*, 2022a.

Lin, H., Huang, Y., Zhang, H., Wu, L., Li, S., Chen, Z., and Li, S. Z. Functional-group-based diffusion for pocket-specific molecule generation and elaboration. *arXiv preprint arXiv:2306.13769*, 2023a.

Lin, H., Wu, L., Xu, Y., Huang, Y., Li, S., Zhao, G., and Li, S. Z. Non-equispaced fourier neural solvers for pdes, 2023b.

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, 2022b.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., 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*, 2022c.

Liu, X., Wu, L., Ye, M., and Liu, Q. Let us build bridges: Understanding and extending diffusion generative models, 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., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *bioRxiv*, 2022. doi: 10.1101/2022.06.06.495043. URL https://www.biorxiv.org/content/early/2022/10/25/2022.06.06.495043.

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):1–20, 2021.

McPartlon, M. and Xu, J. Deep learning for flexible and site-specific protein docking and design. *bioRxiv*, pp. 2023–04, 2023.

Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, 55(14):6582–6594, 2012.

Pei, Q., Gao, K., Wu, L., Zhu, J., Xia, Y., Xie, S., Qin, T., He, K., Liu, T.-Y., and Yan, R. FABind: Fast and accurate protein-ligand binding. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=PnWakgg1RL.

Peluchetti, S. Non-denoising forward-time diffusions. *arxiv*, 2021.

Plainer, M., Toth, M., Dobers, S., Stärk, H., Corso, G., Marquet, C., and Barzilay, R. Diffdock-pocket: Diffusion for pocket-level docking with sidechain flexibility. In *NeurIPS 2023 Workshop on New Frontiers of AI for Drug Discovery and Development*, 2023.

Qiao, Z., Nie, W., Vahdat, A., Miller III, T. F., and Anandkumar, A. State-specific protein-ligand complex structure prediction with a multi-scale deep generative model. *Preprint at arXiv https://doi. org/10.48550/arXiv*, 2209, 2023.

Sahu, D., Rathor, L. S., Dwivedi, S. D., Shah, K., Chauhan, N. S., Singh, M. R., and Singh, D. A review on molecular docking as an interpretative tool for molecular targets in disease management. *ASSAY and Drug Development Technologies*, 22(1):40–50, 2024.

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry*, 49(2): 534–553, 2006.

Shi, Y., Bortoli, V. D., Campbell, A., and Doucet, A. Diffusion schrödinger bridge matching, 2023.

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.

Stärk, H., Ganea, O.-E., Pattanaik, L., Barzilay, R., and Jaakkola, T. Equibind: Geometric deep learning for drug binding structure prediction, 2022.

Sverrisson, F., Feydy, J., Southern, J., Bronstein, M. M., and Correia, B. E. Physics-informed deep neural network for rigid-body protein docking. In *MLDD workshop of ICLR 2022*, 2022.

Taylor, R. D., Jewsbury, P. J., and Essex, J. W. Fds: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *Journal of computational chemistry*, 24(13):1637–1656, 2003.

Vakser, I. A. Protein-protein docking: From interaction to interactome. *Biophysical journal*, 107(8):1785–1793, 2014.

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. Improved protein–ligand docking using gold. *Proteins: Structure, Function, and Bioinformatics*, 52(4):609–623, 2003.

Wu, L., Gong, C., Liu, X., Ye, M., and Liu, Q. Diffusion-based molecule generation with informative prior bridges. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 36533–36545. Curran Associates, Inc., 2022a.

Wu, L., Huang, Y., Lin, H., and Li, S. Z. A survey on protein representation learning: Retrospect and prospect, 2022b.

Wu, L., Huang, Y., Lin, H., Liu, Z., Fan, T., and Li, S. Z. Automated graph self-supervised learning via multi-teacher knowledge distillation, 2022c.

Wu, L., Lin, H., Huang, Y., and Li, S. Z. Knowledge distillation improves graph structure augmentation for graph neural networks. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11815–11827. Curran Associates, Inc., 2022d.

Wu, L., Lin, H., Huang, Y., Fan, T., and Li, S. Z. Extracting low-/high- frequency knowledge from graph neural networks and injecting it into mlps: An effective gnn-to-mlp distillation framework, 2023a.

Wu, L., Lin, H., Huang, Y., and Li, S. Z. Quantifying the knowledge in gnns for reliable distillation into mlps, 2023b.

Wu, L., Lin, H., Liu, Z., Liu, Z., Huang, Y., and Li, S. Z. Homophily-enhanced self-supervision for graph structure learning: Insights and directions. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023c. doi: 10.1109/TNNLS.2023.3257325.

Yan, Y., Zhang, D., Zhou, P., Li, B., and Huang, S.-Y. Hdock: a web server for protein–protein and protein–dna/rna docking based on a hybrid strategy. *Nucleic acids research*, 45(W1):W365–W373, 2017.

Ye, J., McGinnis, S., and Madden, T. L. Blast: improvements for better sequence analysis. *Nucleic acids research*, 34(suppl_2):W6–W9, 2006.

Zhang, O., Wang, T., Weng, G., Jiang, D., Wang, N., Wang, X., Zhao, H., Wu, J., Wang, E., Chen, G., et al. Learning on topological surface and geometric structure for 3d molecular generation. *Nature Computational Science*, 3 (10):849–859, 2023a.

Zhang, Y., Zhang, Z., Zhong, B., Misra, S., and Tang, J. Diffpack: A torsional diffusion model for autoregressive protein side-chain packing. *arXiv preprint arXiv:2306.01794*, 2023b.

Zhang, Z., Xu, M., Jamasb, A., Chenthamarakshan, V., Lozano, A., Das, P., and Tang, J. Protein representation learning by geometric structure pretraining, 2023c.

Zhao, H. and Caflisch, A. Discovery of zap70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. *Bioorganic & medicinal chemistry letters*, 23(20):5721–5726, 2013.

Zheng, J., Li, S., Huang, Y., Gao, Z., Tan, C., Hu, B., Xia, J., Wang, G., and Li, S. Z. Mmdesign: Multi-modality transfer learning for generative protein design, 2023a.

Zheng, J., Wang, G., Huang, Y., Hu, B., Li, S., Tan, C., Fan, X., and Li, S. Z. Lightweight contrastive protein structure-sequence transformation, 2023b.

Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., and Ke, G. Uni-mol: A universal 3d molecular representation learning framework. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=6K2RM6wVqKu.

# A. More related works

**Protein Protein Docking** Protein-protein docking aims to predict the combined structure of two proteins from their individual shapes. This process, similar to molecular docking, typically assumes that proteins do not change shape when they bind, limiting their movements to rotations and translations in three dimensions. Searching-based protein-protein docking methods often start with a basic initialization and refine it gradually (Yan et al., 2017; Chen et al., 2003; De Vries et al., 2010). Furthermore, some use template-based modeling to align a target protein with known structures (Vakser, 2014). Deep learning techniques (Wu et al., 2023a; Lin et al., 2023b; Wu et al., 2023b; Hu et al., 2023; Wu et al., 2022c; Zheng et al., 2023a) in this field are divided into two main types: one aims for immediate predictions, while the other focuses on refining predictions step by step. Research like (Ganea et al., 2021) targets predicting the precise adjustments needed for binding. Another study integrates physical principles into an energy-based model to predict the 3D structure of protein complexes (Sverrisson et al., 2022). Techniques that refine their predictions progressively, such as ALPHAFOLD-MULTIMER (Evans et al., 2021), fold multiple proteins together based on their sequences and related proteins. In parallel with this work, DOCKGPT (McPartlon & Xu, 2023) offers a novel approach for precise and adaptable protein docking.

**Flexible Docking** *Flexible docking* is an open challenge for developing docking methods as it includes much more degrees of freedom than rigid docking. Methods often adopt effective assumptions for restricting the degrees of freedom, such as only considering the flexible sidechain conformations (Alhossary et al., 2015; Taylor et al., 2003; Hartmann et al., 2009). (Zhang et al., 2023b) develop generative models for sidechain packing (that's, without ligands), and (Plainer et al., 2023) explores pocket-specific docking with receptor sidechain flexibility. Different from previous diffusion-based methods for sidechain flexibility, we explore a novel generative framework, incorporate interaction prior for sufficient protein-ligand interaction modeling, and co-model the poses and binding energy. Without sidechain matching and additional training on ESMFold-generated structures, we achieve comparable or better results than (Plainer et al., 2023).

# B. Model Architecture and Training

## B.1. Model Architecture

**Notations.** For each protein-ligand complex, we represent it as a heterogeneous graph with both coarse and fine grain modeling $\mathcal{G} = (\mathcal{V} := \{\mathcal{V}_L, \mathcal{V}_P^{res}, \mathcal{V}_P^{atom}\}, \mathcal{E} := \{\mathcal{E}_{in}, \mathcal{E}_{ex}\})$. Here, the components $\mathcal{V}_L, \mathcal{V}_P^{atom}$ correspond to the nodes (i.e., the heavy atoms) of the ligand and protein (atom-level). $\mathcal{V}_P^{res}$ represent the nodes (i.e., the residues) of the protein (residue-level); $\mathcal{E}_{in}, \mathcal{E}_{ex}$ separately contain internal edges within each component and external edges across components. To be specific, each node in $\mathcal{V}$, i.e., $v_i = (h_i, x_i)$ is represented as a node embedding vector $h_i$ initialized with trainable type embedding vector and ESM-2 embedding (Lin et al., 2022c) following (Corso et al., 2023), and a coordinate vector $x_i \in \mathbb{R}^3$ of the corresponding (heavy) atoms (atom-level) or the alpha carbon of the residue backbone (residue-level). To be aware of the general gesture of ligand for better generation of the binding pose, we add a global node that connects to all atoms in the component coordinated at the mean of all coordinates.

Each edge in $\mathcal{E}$, i.e., $e_{ij} = (h_{ij}, x_{ij})$ consists of an edge embedding of node $(i, j)$ initialized with radial basis embedding of edge length and edge vector $x_{ij} = x_j - x_i$. For explicit and accurate modeling of protein-ligand interactions, we separate internal and external interactions ($\mathcal{E}_{in}$ and $\mathcal{E}_{ex}$ respectively) because of their different role in binding and distance scales. Furthermore, to model interactions in a fine-grained manner, We connect nodes using cutoffs dependent on the type of nodes they are connecting and assign the edges distinct edge types corresponding to different graph convolution kernels following (Zhang et al., 2023c; Kong et al., 2023; Corso et al., 2023). We further denote the pocket-ligand subgraph as $\mathcal{G}_{p*} = (\mathcal{V} := \{\mathcal{V}_L, \mathcal{V}_{p*}^{res}, \mathcal{V}_{p*}^{atom}\}, \mathcal{E} := \{\mathcal{E}_{in}, \mathcal{E}_{ex}\})$ which is the input of the docking model.

**Architecture.** We construct a heterogeneous graph with atoms and residue levels to reason the pocket-ligand interaction in a more fine-grained manner (Zhang et al., 2023c). The edge constructions have considered various interaction types, including internal interactions ( atom-atom interactions, chemical bone interactions, residue-residue spatial and sequential interactions, atom-residue subordination relationship, etc.) and external interactions (pocket-ligand atom interactions, residue-atom interactions, framework-pocket interactions, etc.). We assign different graph convolution kernels to each edge type and model every interaction separately. These relation-aware schemes can enhance the modeling ability for accurate pocket-ligand interactions.

The output of the score model must be in the tangent space $T_{\mathbf{r}}\mathbb{T}_3 \oplus T_R SO(3) \oplus T_{\boldsymbol{\theta}} SO(2)^m$ and the predicted energy must be in $SE(3)$-invariant. The space $T_{\boldsymbol{r}}\mathbb{T}_3$ and $T_R SO(3)$ represent the translation and rotation (Euler) vectors which are both $SE(3)$-equivariant and $T_{\boldsymbol{\theta}} SO(2)^m$ corresponds to scores on $SE(3)$-invariant quantities (torsion angles). To achieve such

desire, we apply a $SE(3)$-equivariant convolution network(6 layers of graph convolutions on the heterogeneous graph implemented with e3nn library (Geiger & Smidt, 2022; Corso et al., 2023)) as the shared encoder with equivariant and invariant pooling as heads as well as pseudo torque convolution (Jing et al., 2022) for predicting torsional updates.

### B.2. Model training and implementation details

Following (Corso et al., 2023), we train and evaluate all the models on PDBBind (Liu et al., 2017) based on time-split dataset partition. We also adopt a conformer matching procedure described in (Jing et al., 2022) for eliminating the distributional shift between RDKit-initialization used in inference and ground-truth ligand pose in the dataset. We summarize the training as algorithmic description as in Algorithm 1.

**Algorithm 1** Learning diffusion generative models

**Input**: Training pairs $\{(x^*, y\}$ where $x^*$ is the ground truth ligand pose and $y$ is the ground truth protein structure, RDKit preditions $\{c\}$, variance for each geometry $\{\sigma_{\mathfrak{g}}\}$, prior interaction energy $E(x, y)$ and $\mathbb{Q}^x_{\mathfrak{g}}$ the bridge in Eq. 6, and energy-to-geometry mapping $F$ and a diffusion model $\mathbb{P}^\theta$.
**Output**: Parameters $\theta$ of $\mathbb{P}_\theta$

Randomly initialize the parameters of $\mathbb{P}_\theta$.
**for** $c, x^*, y \in \{(x^*, y, c)\}$ **do**
  Let $x_0 \leftarrow conformer_align(x^*, c)$;
  Sample $t \in Uni([0, 1])$, $chi_idsc \in 0, 1, 2, 3$;
  Sample $\{\Delta\mathfrak{g}\}$ from diffusion kernels $\{p^{\mathfrak{g}}_t(\cdot|\sigma_{\mathfrak{g}})\}$;
  Compute $(x_t, y_t) \leftarrow Apply(\{\Delta\mathfrak{g}\}, x_0, y, sc)$ and $f^{\mathfrak{g}}_t = F(E(x_t, y_t))$;
  Predict scores;
  Take optimization step on the sum of denoising score matching loss Eq. 4 on each geometry;
**end for**
**return** Diffusion Bridge $\mathbb{P}_\theta$.

**training detail.** We use Adam as optimizer with learning_rate= 0.001 and exponential moving average of the weights during training, which we will use in inference. we update the moving average after every optimization step with a decay factor of 0.999. The batch size is 64. We run inference with 20 denoising steps on 500 validation complexes every 10 epochs and use the set of weights with the highest percentage of RMSDs less than 2Å as the final score model. All baselines and our approach are implemented using the PyTorch 1.6.0 library with Intel(R) Xeon(R)Gold6240R@2.40GHz CPU and NVIDIA A100 GPU. We train our score model for 600 epochs (around 7 days). As we co-model the binding energy and poses, we don't need additional training for the confidence model as in (Corso et al., 2023), but the confidence model can still be adopted in our framework for ranking. For inference, only a single GPU is required but we also implement parallel inference in multi-gpu settings for large scale screening task like cross-dock (around 10000 complexes, 10 poses per complexes and 20 time steps).

**Runtime** Similar to previous methods, we exclude the data preprocessing time and perform runtime analysis on single cpu core (plus additional gpus for methods that can utilize gpus). Preprocessing mainly consists of a forward pass of ESM2 to generate the protein language model embeddings, RDKit's conformer generation, and the conversion of the protein into a radius graph. We measured the inference time when running on the same device for training.

## C. Experimental setup

To evaluate the performance of our method, we design a systematic benchmark mimicking the flexible and realistic setting in drug discovery pipelines, including redocking, unbound(apo)-structure docking (including predicted structures and experimental structures), and cross-dock (docking to bound structure with another ligand, which is very challenging but common in drug discovery) experiments in the PDBBind dataset. Re-Dock has shown competitive results across all tasks in various metrics assessing conformational plausibility. This suggests our model can handle the flexibility of both pockets and ligands and has the potential to be helpful in realistic drug discovery pipelines.

### C.1. Data Curation

**Crystal apo structure for PDBBind time split test set.** The APO structures are retrieved from the Protein Data Bank (PDB) (Burley et al., 2021) by first extracting the sequence of HOLO structures and conducting a BLAST (Ye et al., 2006) search against the PDB database. Each hit protein is then structurally aligned to the holo-structure using PyMOL (DeLano et al., 2002), focusing on the superposition of corresponding $C_\alpha$ atoms of amino acid residues. Post alignment, structures are assessed for quality and relevance similar to ApoBind (Aggarwal et al., 2021): those with a backbone $C_\alpha$ Root Mean Square Deviation (RMSD) exceeding 15 Å, or those showing less than 80% sequence identity or coverage compared to the full protein sequence, are rejected. Additionally, any hit structure with ligands located within 4 Å of any atoms in the crystal structure pose of the complex is also excluded.

**Crystal CrossDock structures.** The targets are from DUD-E (Mysinger et al., 2012), a dataset designed for the unbiased virtual screening task. Seven targets are selected, AKT1, AMPC, CXCR4, GCR, HIVPR, HIVRT, and KIF11, which are representatives of kinase, other enzymes, G protein-coupled receptors, nuclear receptors, protease, other enzymes, and miscellaneous classes. Within each target, we retrieve different protein-ligand structures from PDB using a similar method and conditions of the construction of crystal apo time split test set.

For search-based method, we use their official software suits and for deep-learning-based baselines, we use their official implementation and weights. To develop diffdock (pocket), we adopt the same pocket truncation method as our RE-DOCK and full-atom graph representation as it used for their confidence model. Other settings keep the same with the original diffdock.

## D. More details about the bridge

The space of the pocket-ligand poses $\mathbb{R}^{3(m+n)}$ (m and n are the number of pocket sidechain and ligand atoms) is huge and encompasses far more degrees of freedom than are relevant in molecular docking. The complex flexibility lies almost entirely in the torsion angles at rotatable bonds (Corso et al., 2023); thus, we incorporate geometric prior, which is known in advance (e.g., fixed bond lengths, angles, and essentially rigid small rings) with a seed (randomized only in torsion angles) pocket sidechain and ligand conformations $\mathbf{C}$ in isolation, and model the pocket-ligand complex pose in an $(m+6)$-dimensional submanifold $\mathcal{M}_C$, where $m$ is the number of rotatable bonds and six additional degrees of freedom come from rototranslations relative to the fixed protein backbone.

While diffusion generative models have been applied to molecular docking, existing approaches are ill-suited for flexible and realistic docking scenarios, where we need to co-model the flexibility of pocket sidechains and ligands as well as their interactions. To develop Re-Dock, we recognize the dynamic and interactive nature of docking with induced fit; thus, we use the diffusion bridge model to inject physical priors of molecular interactions into the generative process. To simplify the modeling process and incorporate geometric prior (e.g. fixed bond lengths, bond angles between atoms), we follow the successful experience in the field of conformational generation and develop geometry-based generative models.

We construct such a bridge with prior of the atom interaction potentials regarding how the diffusion process should look like for generating each given data point over the product space of geometries: global rotation/translation of ligands and local molecular torsion/residue sidechain angles.

As the induced-fit process is driven by the inherent interaction energies and generative docking needs a confidence model for ranking predicted poses (Corso et al., 2023), we include learnable interaction energy with an additional energy head to co-model the interaction energy and conformation.

We use prior energy terms in Amber (Case et al., 2021), including: 1. The Lennard-Jones (LJ) energy $E_{LJ}(x) = \sum_{i \neq j} e(\|x_i^r - x_j^r\|)$ and $e(\ell) = (\sigma/\ell)^{12} - 2(\sigma/\ell)^6$. The parameter $\sigma$ is an approximation for average nucleus distance. 2. The nuclei-nuclei repulsion (Coulomb) electromagnetic potential energy is $E_{Coulomb} = \kappa \sum_{ij} q(\hat{x}_i^h) q(\hat{x}_j^h) / \|x_i^r - x_j^r\|$, where $\kappa$ is Coulomb constant and $q(r)$ denotes the point charge of atom of type $r$, which depends on the number of protons. We use these energies based on the same threhold with the $E_{clash}$ in Equation. 10 as the ligand and sidechains are close enough.

As the *energy-to-geometry mappp* is derived from rigid-body mechanics, it is naturally equivariant to roto-translations (i.e. rotations and translations). As the calculation of prior drift on torsion manifold $SO(2)^M$, where $M = m_{lig} + m_{sc}$ is the number of rotatable bonds, is complex and requires extensive computations, we omit it for simplicity. The torsion force can be described as the difference between the torques applied on each side of the rotatable bonds,but its adaptation with

$SO(2)^M$ manifold and autoregressive sidechain torsion update is non-trivial and we leave it as future work.

## E. More Results

We provide generlized results with predicted pocket with p2rank (Krivák & Hoksza, 2018) in table 6

| | Holo Crystal Proteins | | | | | | Apo ESMFold Proteins | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Top-1 RMSD | | | Top-5 RMSD | | | Top-1 RMSD | | | Top-5 RMSD | | | Average |
| Method | %<2 | %<5 | Med. | %<2 | %<5 | Med. | %<2 | %<5 | Med. | %<2 | %<5 | Med. | Runtime (s) |
| GNINA | 34.6 | 62.6 | 3.3 | 46.8 | 78.1 | 2.1 | 9.9 | 31.7 | 6.7 | 18.1 | 53.5 | 4.6 | 251 |
| SMINA | 21.5 | 50.1 | 5.0 | 37.2 | 70.8 | 2.8 | 4.9 | 19.2 | 7.3 | 9.6 | 43.4 | 5.4 | 246 |
| LeDock | 24.0 | 48.6 | 5.1 | 39.4 | 69.8 | 2.9 | 10.9 | 36.7 | 6.2 | 19.5 | 59.0 | 4.2 | 7 |
| SurfLex | 6.1 | 31.9 | 6.7 | 10.6 | 53.1 | 4.7 | 1.4 | 15.6 | 7.4 | 2.6 | 30.1 | 6.1 | 5 |
| QVina | 21.3 | 50.9 | 4.9 | 36.8 | 71.8 | 2.9 | 5.0 | 19.5 | 7.3 | 9.9 | 42.9 | 5.4 | 2 |
| GNINA-Flex | 29.4 | 63.7 | 3.5 | 45.7 | 84.9 | 2.2 | 10.9 | 36.7 | 6.2 | 19.5 | 59.0 | 4.2 | 1072 |
| SMINA-Flex | 20.5 | 46.2 | 5.3 | 34.4 | 73.3 | 2.9 | 5.8 | 18.4 | 7.1 | 11.1 | 42.0 | 5.4 | 963 |
| FlexPose | 42.0 | 79.3 | 2.2 | - | - | - | 34.8 | **79.9** | 2.8 | - | - | - | 12 |
| DiffDock(pocket, 40) | 51.8 | 75.3 | 2.0 | 60.7 | 79.2 | 1.9 | 37.8 | 72.4 | 2.6 | 38.1 | 73.4 | 2.5 | 61 |
| **ReDock (10)** | 51.9 | 77.4 | 2.0 | 62.2 | 83.6 | 1.7 | 39.0 | 74.8 | 2.5 | 39.0 | 74.8 | 2.5 | 15 |
| **ReDock (40)** | **53.9** | **80.3** | **1.8** | **65.0** | **86.7** | **1.4** | **42.9** | 76.4 | **2.4** | **45.6** | **78.2** | **2.2** | 58 |
| **ReDock (p2rank, 40)** | 40.2 | 65.7 | 3.1 | 45.6 | 77.8 | 2.2 | 22.8 | 56.7 | 4.9 | 32.8 | 65.2 | 3.3 | 58 |

*Table 6.* Performance of flexible redocking on the PBDBind test set (bound, holo-crystal proteins) and its corresponding ESMFold predicted Apo structures (unbound). The best metrics are marked by **bold**. In parenthesis, we specify the number of poses sampled from the generative model. It is worth noting that only our `Re-Dock` has no access to the holo-crystal proteins. p2rank denotes we use the predicted pocket center with p2rank and a pocket radius of 15 Å