Latent Diffusion for Ligand-Protein Docking

Le-Tung Hsieh, Fengzhi Guo, Yu-Sheng Chen
Texas A&M University
{lthsieh, fengzh_g, chenyus0609}@tamu.edu

Abstract

Ligand-target protein docking is pivotal in drug discovery, yet previous methods either face computational complexity or struggle to capture molecular dynamics effectively. Recently, there has been a novel idea for de novo design which leverages latent 3D space representations with diffusion models attempting to overcome these challenges. Latent representation integrates additional information, allowing a wider, richer scope for exploration to enhance predictive binding performance. This report presents an innovative approach for docking inspired by recent advancements, which simulates dynamic ligand-protein interactions. It explores latent conformational space, aiding in energetically guiding favorable binding configurations. The model tries to predict binding affinities and kinetics by simulating ligand diffusion towards binding sites. This report consists of experimental evaluations to show our work results, obstacles, explanations, and future works.

1 Introduction

1.1 Problem

1.1.1 General Topic

Docking is an approach akin to solve a puzzle in molecular biology and drug design, involves fitting a ligand into a protein to find the best arrangement. This process is crucial in rational drug design, offering an efficient and economical approach for discovering lead molecules targeted at known protein binding pockets. Conventional docking methods typically use scoring functions to gauge the accuracy of a given structure, coupled with optimization algorithms tasked with locating the global maximum of these scores. However, it remains challenging and computationally intensive due to the vast synthetically feasible space(1) and the high degrees of freedom for binding poses(2). While previous molecular generative models have focused on molecular string or graph representations, they often overlook 3D spatial interactions essential for target-aware molecule generation. Recent advancements in structural biology and protein structure prediction have led to the availability of more structural data, unlocking new opportunities for machine learning algorithms to design drugs directly within 3D binding complexes(3).

1.1.2 Current Methods

Recently, a number of novel generative models have surfaced, aimed at exploring ligand poses in molecular docking and enabling target-aware generation. TargetDiff(4), built on recent advancements in probabilistic diffusion models and equivariant neural networks, accounting for protein context while maintaining translational and rotational invariance. It presents protein binding pockets and small molecules as atom point sets in 3D space, with a joint generative process learned using a SE(3)-equivariant graph neural network. DIFFDOCK(5), a diffusion generative model (DGM), involves establishing a diffusion process across the various degrees of freedom inherent docking, encompassing the ligand's positioning relative to the protein, its orientation within the pocket, and the torsion angles characterizing its conformation. DIFFDOCK generates pose samples by executing the learned diffusion process, which progressively refines an initial, noisy prior distribution of ligand poses into the modeled distribution. Lately, You et al.(2024)(6) proposed a SOTA approach which improves 3D graph diffusion models by introducing a lower-dimensional latent space that strikes a balance between reconstruction accuracy and data complexity. The approach employs a pre-trained 3D graph autoencoder (AE) to map non-Euclidean structures into latent embeddings and a diffusion generative model (DGM) to capture data distributions within this latent space. By integrating SE(3) invariant or equivariant conditions, it enhances the conditional generation capacity of DGM, ensuring symmetry preservation while optimizing reconstruction quality relative to data dimensionality.

1.1.3 Remaining Gaps

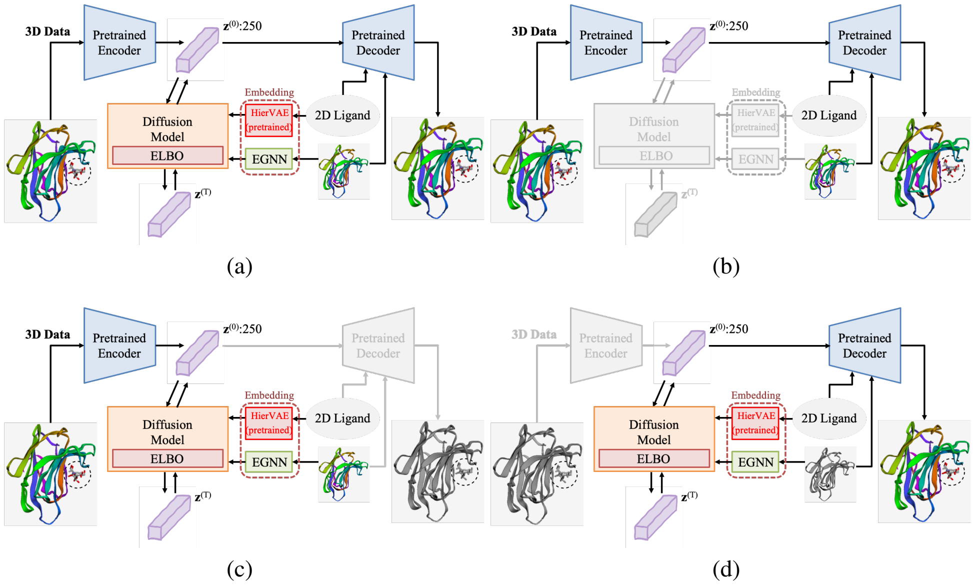
In contrast to the search-based approach, DIFFDOCK exhibits superior precision and efficiency. However, it is imperative to critically reevaluate the foundational assumption positing the fixed nature of bond lengths, angles, and small ring structures within the ligand, while ascribing flexibility primarily to torsion angles at rotatable bonds. Within this paradigm, the three-dimensional conformation of the ligand is anticipated through the utilization of RDKit(7), with bond lengths and angles held invariant. The conceptualization of the solution space adopts a framework of Cartesian product, encompassing rotations, translations, and torsions exclusively. Nevertheless, it is crucial to acknowledge that bond lengths and angles adhere to a probabilistic distribution rather than absolute fixed values. This disjunction between prescribed parameters and empirical actuality engenders a fundamental impediment, rendering the realization of an optimal solution unattainable within the predefined product space. Additionally, although navigating the docking challenge within a $6 + m$ dimensional submanifold might seem simpler compared to the original $3n$ dimensional space, the increase in torsion angles corresponding to the complexity of the ligand amplifies the vastness of the solution space by m, consequently requiring more computational resources.

1.1.4 Proposed Method Overview

We discover that the latent space serves as a versatile representation of raw data, while the product space can be seen as a specialized version of the latent space, governed by physical laws. To enhance the quality of latent representation, we draw inspiration from the Latent 3D Diffusion Model for innovating ligand design. In this approach, a pretrained autoencoder effectively maps 2D and 3D ligands onto the latent space. For novel ligand design, we sample a noisy latent vector at time T and leverage the diffusion model conditioned on the protein to obtain a refined latent vector. Upon decoding, we obtain novel ligands with both 2D and 3D representations. Inspired by this methodology, we construct a neural network structure tailored for the docking problem.

2 Methods

Fig. 1(a) illustrates the overall framework, which can be segmented into three main pipelines. Initially, as shown in Fig. 1(b), the autoencoder for 3D ligand is trained (6). Inputting the 3D ligand, the encoder maps it to the latent space. Subsequently, during the decoding phase, the latent vector with the 2D ligand and 3D protein coordinates is utilized to reconstruct the 3D ligand. Next, the diffusion model training process is illustrated in Fig. 1(c). Employing the pretrained encoder, we produce the latent vectors as the groundtruth values. In the typical training sequence for the diffusion model, noise is progressively added to the data from the initial time (time 0) to the final time (time T), known as the forward process. Subsequently, the diffusion model is employed to predict the noise in the reverse denoising phase. The loss is calculated as the distance between the predicted noise and the ground truth noise. In this docking scenario, the latent vector represents the 3D ligand, and it is understood that there is a correlation between the latent vector and the 2D ligand and the 3D protein coordinates for docking purposes. Thus, the diffusion model is adapted to depend on the 2D ligand and 3D protein coordinates. Additionally, since the 2D ligand and 3D protein cannot be directly fed into the diffusion model, they are embedded using HierVAE (8) and EGNN (9), respectively. Throughout the training phase, the HierVAE is pretrained to output 2D ligand embeddings while the noise prediction model and EGNN are trained concurrently. In the final phase shown in Fig. 1(d), a latent noised vector is sampled at time T, and the diffusion model, conditioned on the specified 2D ligand and 3D protein, is used to produce the latent denoised vector at time 0. Following the decoding process, the docking results are obtained.



(a)

(b)

(c)

(d)

Figure 1: Latent diffusion model. (a) Overall architecture. (b) Autoencoder training pipeline. (c) Diffusion model training pipeline. (d) Sampling and docking pipeline. $(\mathbf{z}^{(i)} : 250)$ represents the latent vector for 3D ligand at time i with length equal to 250.

3 Experiment

3.1 Design

We follow the work of the pre-trained autoencoder in (6) and train the diffusion model and protein encoder. The model is trained and test on the dataset of CrossDocked (100K complexes) (10; 4). To evaluate the docking results, we choose the Vinardo (Vina) scoring function, which is commonly used in molecular docking to evaluate the binding affinity between a protein and a ligand. It employs a combination of empirical and knowledge-based terms to estimate the binding free energy.

3.2 Results

Table 1: Docking Results. True (T) and False (F).

#	Trained Ligand?	2D Decoder?	Vina Score	Vina Min	Vina Dock	Rate
1	T	T	-4.836	-4.608	-8.011	14.7%
2	F	T	-4.185	-5.483	-9.275	1.3%
3	F	F	N/A	N/A	N/A	0.0%

The docking results are shown in Table 1. The ideal docking pipeline is shown in Fig. 1(d). The inputs are the 2D ligand and protein in the test dataset. Considering the previous work (6) only validate the generalization ability with protein as input, we still need to analyze the sampling results with both proteins and topology of ligands as input. For the purpose of comparison, we choose to sample docking results with 2D ligand in the training set and replace 2D groundtruth ligand with the output by inputting the groundtruth ones into the autoencoder. Besides, we also list the successful rate of the successful docking among all samples. For all the three experiments settings, the proteins come from the test dataset. Comparing with settings #1 and #2, without using the 2D trained ligand as inputs, the successful rate drops quickly. Next, by using 2D groundtruth smiles files rather than the ones passing the autoencoder in setting #3, no samples can be successfully docked because explicit valence for parts of atoms is greater than permitted value.

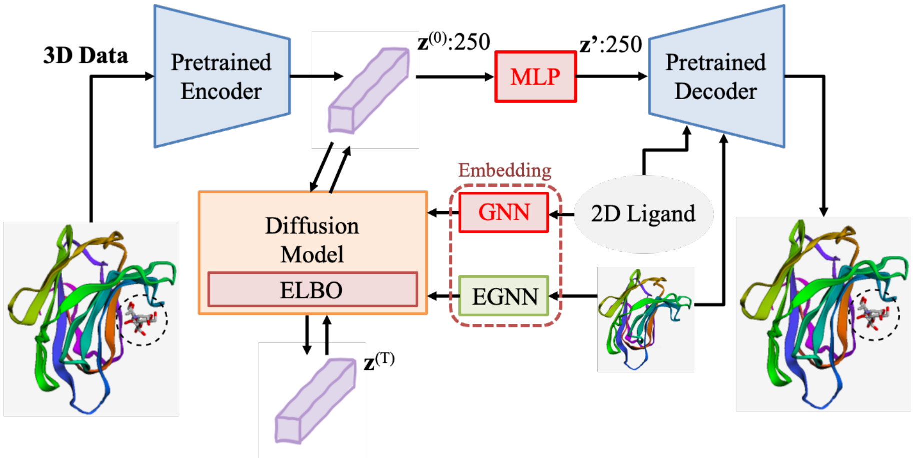


Figure 2: Overall architecture for future works.

3.3 Experiment Interpretations

According to the experimental results, it seems that although the "Latent-Diffusion" model architecture is good at generating novel ligands for a given protein (de novo design), it had failed to reconstruct the protein-ligand docking pair given restrictions on both protein and ligands.

Notice that if we are using "generated 2D graphs" (no diffusion involved, only encoder-decoder involved, which will generate 2D graphs that are very close, yet may not be exactly the same as ground truth), we have a probability of successful docking output, while if we are using ground truth 2D graphs, we failed to reconstruct the protein-ligand docking pair.

The following are some possible explanations for the results.

3.3.1 Uncovered Distribution

It is possible that the poor docking results are due to a possibility that the distribution of training and test set are too different from each other. Even though the model is supposed to generalize to unseen instances, we do have to keep in mind that unseen features may lead to poor performance.

3.3.2 Pre-trained Geometric Decoder

Another possibility may be caused by the fact that we are using pre-trained geometric-decoder weights that may be trained from the output of the topological decoder. However, when we are doing docking (our modification), we have to fix the 2D information of the ligand (using ground truth smiles of the ligand), while the geometric decoder may be pre-trained on the "synthesized smiles". This means the distribution that the geometric decoder learned from may be different from the ground truth, which eventually led to poor docking results.

4 Future Works

4.1 Energy Guidance Model

Fig. 2 illustrates the concept of our future workflow pipeline and diagram. We endeavored to improve our proposed model's performance by integrating physics-informed methods. Understanding that docking efficiency improves as the energy decreases post-ligand attachment to the target protein, we sought inspiration from the presentation of classmates, Ali and Amir, to introduce a new module before the decoder. Drawing from their insights, we propose integrating a Multilayer Perceptron (MLP) to optimize energy loss. This loss metric is basically defined as the disparity between the outcome after the decoder and the initial docking conditions. During training, the decoder remains frozen while the gradient of the loss is propagated backward to refine the parameters of this MLP block. The goal is to produce optimal ligand docking configurations with minimized and stable energy levels.

4.2 GNN Model for 2D Ligand Embeddings

We tried to replace HierVAE with Graph Neural-Networks (GNNs) as the encoder of 2D ligand features (in graphs). Attempting to fully replace the topological encoder-decoder structure in the original model, so that it will not only cope better with our docking problem, but also reduce the computational time, making the model more efficient. After converting the 2D ligands from smiles (strings) to graphs, they are sent to a GNN-based embedding block, which will output a tensor with dimension 250, which we shall consider it as being in the "same latent space" as the original version. For training, we let the GNN learn to map graphs of 2D ligands to embeddings which are collected from the original HierVAE encoder, so that we can replace the role of it. For inference mode, we let ground truth ligands (which should be fixed under the docking scenarios) pass through the GNN block and act as a condition for docking diffusion. (See Fig. 2). So far, we have trained a GNN that gives a MSE Loss of 30.7 on training dataset, which is far from satisfaction. Therefore, we have decided not to integrate this structure with the diffusion model in the aforementioned experiments (which should be retrained upon this new setting), since poor performance can be expected. Nonetheless, this can still be a good starting point for improving the whole structure.

5 Conclusions

5.1 Current Works

Currently, we have tried to solve the ligand-protein docking problem based on the work of the Latent 3D Graph Diffusion structure. The training process includes the 3D ligand autoencoder and diffusion model training while in the inference process, we need to input the 2D ligand and protein as the conditions into the diffusion model to generate the 3D ligand as the docking results. The experiments results show the potential of the docking capability. However, because of the aforementioned uncovered distribution and the problem of the pre-trainined geometric decoder, the large explicit valence leads to failure of reasonable 3D ligand for docking. Besides, we also did preliminary exploration on the energy guidance model and smaller GNN model for end-to-end training. More concerns are listed in limitations.

5.2 Limitations

Although putting the diffusion part in the latent space is a smart move, that might give potential to integrating more features of the ligands themselves and allowing a broader scope for diffusing, a big downside of this method is that it is impossible, or extremely difficult to integrate physics laws in the scoring function of the diffusion model. When attempting to develop and implement the energy guidance model, we encountered a challenge regarding the integration of RDKit for calculating and training newly inserted blocks into the pipeline. Although RDKit effectively computes the energy difference between conformers, it is limited to providing specific values without capability to retroactively feed this information back into the MLP for model training. To address this limitation, it is needed to incorporate a cutting-edge model capable of storing and backwardly propagating values to achieve optimized results.

5.3 Future Directions

As stated in section 4, the future directions of this project are integrating energy guidance and replacement of encoder for 2D ligands. Another direction can be using Graph Attention Networks (GATs) for the GNN block in section 4, for this is a newer and more powerful kind of networks than simple GNNs.

6 Contributions

All authors contributed equally to this work.

F.G., L.H., Y.C. designed the whole experiment and wrote the report.

F.G. adapted the Latent 3D model to address the docking problem and conducted the relevant experiments. L.H. delved into the energy guidance related papers and performed experiments of the energy loss. Y.C. delved into the Graph-based Neural Networks and performed experiments of the GNN embedding of 2D graphs of the ligands.

7 Acknowledgements

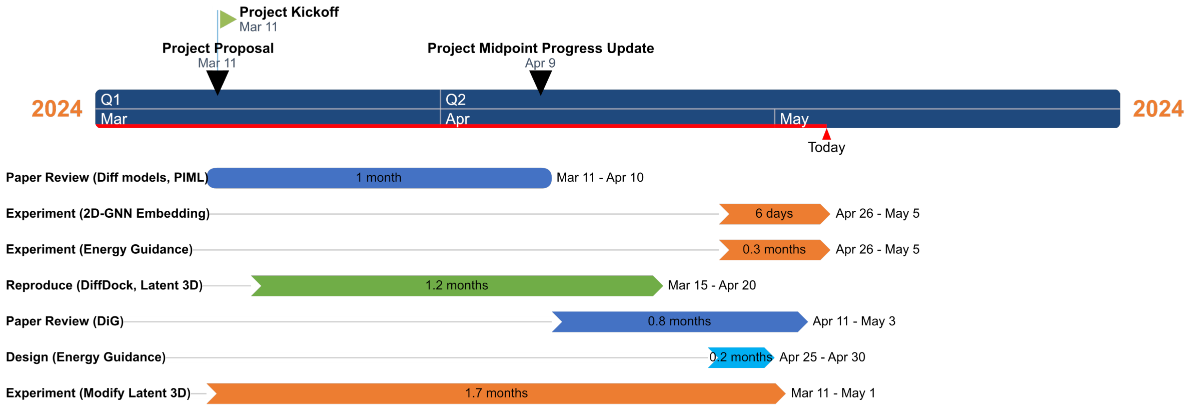
It is impossible for us to get to this far without the help of the following people. We deeply appreciate: Professor Yang Shen for discussing various innovative and smart ideas for the project. The first author of Latent 3D Diffusion Model, Yuning You for assisting us tracing the codes. pointing out directions in such a large project. Our classmates, Ali and Amir, for providing such brilliant idea for how to integrating energy into the architecture as well as the great discussions between us.

8 Appendix

8.1 Availability

The code can be found in the *dock* branch of the repo as below.
Github Repo: `https://github.com/percool/LDM-3DG.git`

8.2 Execution Gantt Chart



References

[1] Ragoza, M., T. Masuda, D. Koes. Generating 3d molecules conditional on receptor binding sites with deep generative models. *Chemical Science*, 13, 2022.

[2] Hawkins, P. C. D. Conformation generation: The state of the art. *Journal of chemical information and modeling*, 57(8):1747–1756, 2017.

[3] Simm, G., R. Pinsler, J. Hernández-Lobato. Reinforcement learning for molecular design guided by quantum mechanics. 2020.

[4] Guan, J., W. W. Qian, X. Peng, et al. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

[5] Corso, G., H. Stärk, B. Jing, et al. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

[6] You, Y., R. Zhou, J. Park, et al. Latent 3d graph diffusion. In *The Twelfth International Conference on Learning Representations*. 2023.

[7] Landrum, G. Rdkit: Open-source cheminformatics.

[8] Jin, W., R. Barzilay, T. Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, pages 4839–4848. PMLR, 2020.

[9] Stärk, H., O. Ganea, L. Pattanaik, et al. Equibind: Geometric deep learning for drug binding structure prediction. In *International conference on machine learning*, pages 20503–20521. PMLR, 2022.

[10] Francoeur, P. G., T. Masuda, J. Sunseri, et al. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of chemical information and modeling*, 60(9):4200–4215, 2020.