# A IMPLEMENTATION DETAILS

In this section, we will provide more implementation details of our methods. Though some contents, such as the SE(3)-equivariant layer and the loss function, can be referred to Guan et al. (2023b), we still include them here to make our paper more self-containing.

## A.1 FEATURIZATION

We follow the decomposition algorithm proposed by Guan et al. (2023b) to decompose ligand molecules into arms and a scaffold. We define the part of proteins that lies within 10Å of any atom of an arm as its corresponding subpocket.

Following DecompDiff (Guan et al., 2023b), we represent each protein atom with the following features: one-hot element indicator (H, C, N, O, S, Se), one-hot amino acid type indicator (20 dimension), one-dim flag indicating whether the atom is a backbone atom, and one-hot arm/scaffold region indicator. If the distance between the protein atom and any arm center is within 10Å, the protein atom will be labeled as belonging to an arm region and otherwise a scaffold region. The ligand atom is represented with following features: one-hot element indicator (C, N, O, F, P, S, Cl) and one-hot arm/scaffold indicator. Different from DecompDiff, the atom features are enhanced by concatenating an SE(3)-invariant feature of arms and their corresponding subpockets encoded by the condition encoder after the orignal features.

Two graphs are constructed for message passing in the protein-ligand complex: a k-nearest neighbors graph $\mathcal{G}_K$ upon ligand atoms and protein atoms (we choose $k = 32$ in all experiments) and a fully-connected graph $\mathcal{G}_L$ upon ligand atoms. The edge features are the outer products of distance embedding and edge type. The distance embedding is obtained by expanding distance with radial basis functions (RBF) located at 20 centers between 0Å and 10Å. The edge type is a 4-dim one-hot vector indicating the edge is between ligand atoms, protein atoms, ligand-protein atoms or protein-ligand atoms. In the ligand graph, the ligand bond is represented with a one-hot bond type vector (non-bond, single, double, triple, aromatic), an additional feature indicating whether or not two ligand atoms are from the same arm/scaffold.

## A.2 MODEL DETAILS

The controllable and decomposed diffusion model consist of two parts: a condition encoder and a diffusion-based decoder. The building block is an SE(3)-equivariant layer that is composed of three layers: atom update layer, bond update layer, and position update layer.

We denote the protein pocket as $\mathcal{P} = \{(\mathbf{x}_i^{\mathcal{P}}, \mathrm{v}_i^{\mathcal{P}})\}_{i\in\{1,\dots,N_{\mathcal{P}}\}}$ and the ligand molecule as $\mathcal{M} = \{(\boldsymbol{x}_i, \boldsymbol{v}_i, \boldsymbol{b}_{ij})\}_{i,j\in\{1,\dots,N_{\mathcal{M}}\}}$, where $\boldsymbol{x}$ is the atom position, $\boldsymbol{v}$ is the atom type, and $\boldsymbol{b}_{ij}$ is the chemical bond type between the atom $i$ and the atom $j$. For brevity, we omit the superscript $\mathcal{P}$ or $\mathcal{M}$ in the following. We use $\mathbf{h}_i$ to denote the the SE(3)-invariant hidden state of $i$-th atom, $\mathbf{x}_i$ to denote the $i$-th atom's coordinate, which is SE(3)-equivariant, and $\mathbf{e}_{ij}$ to denote the hidden state of the edge between the $i$-th atom and the $j$-th atom. They can be obtained as we described in the previous subsection. And we use $t$ to denote the time embedding as that in Ho et al. (2020).

**Atom Update Layer** We denote the atom update layer as $\phi_a := \{\phi_{a1}, \phi_{a2}, \phi_{a3}, \phi_{a4}\}$.

We first use the atom update layer $\phi_{a1}$ to model protein-ligand interation as follows:

$$\Delta\mathbf{h}_{K,i} \leftarrow \sum_{j\in\mathcal{N}_K(i)} \phi_{a1}(\mathbf{h}_i, \mathbf{h}_j, ||\mathbf{x}_i - \mathbf{x}_j||, \mathbf{e}_{ij}, t), \tag{3}$$

where $\mathcal{N}_K(i)$ is the set of neighbors of the $i$-th atom in the protein-ligand complex graph $\mathcal{G}_K$.

We then further use the atom update layer $\phi_{a2}$ and $\phi_{a3}$ to model the interaction inside the ligand as follows:

$$\mathbf{m}_{ij} \leftarrow \phi_{a2}(||\mathbf{x}_i - \mathbf{x}_j||, \mathbf{e}_{ij}), \tag{4}$$

$$\Delta\mathbf{h}_{L,i} \leftarrow \sum_{j \in \mathcal{N}_L(i)} \phi_{a3}(\mathbf{h}_i, \mathbf{h}_j, \mathbf{m}_{ji}, t), \tag{5}$$

where $\mathcal{N}_L(i)$ represents the set of neighbors of the $i$-th atom in the ligand graph $\mathcal{G}_L$. Finally, we update the hidden state of atoms by the atom update layer $\phi_{a4}$ as follows:

$$\mathbf{h}_i \leftarrow \mathbf{h}_i + \phi_{a4}(\Delta\mathbf{h}_{K,i} + \Delta\mathbf{h}_{L,i}). \tag{6}$$

**Bond Update Layer** We update the hidden states of the edges by the bond update layer $\phi_b$ as follows:

$$\mathbf{e}_{ij} \leftarrow \sum_{k \in \mathcal{N}_L(i) \setminus \{j\}} \phi_b(\mathbf{h}_i, \mathbf{h}_j, \mathbf{h}_k, \mathbf{m}_{kj}, \mathbf{m}_{ji}, t). \tag{7}$$

**Position Update Layer** The atom positions are updated by the position update layer $\phi_p := \{\phi_{p1}, \phi_{p2}\}$ as follows:

$$\Delta\mathbf{x}_{K,i} \leftarrow \sum_{j \in \mathcal{N}_K(i)} (\mathbf{x}_j - \mathbf{x}_i)\phi_{p1}(\mathbf{h}_i, \mathbf{h}_j, ||\mathbf{x}_i - \mathbf{x}_j||, t), \tag{8}$$

$$\Delta\mathbf{x}_{L,i} \leftarrow \sum_{j \in \mathcal{N}_L(i)} (\mathbf{x}_j - \mathbf{x}_i)\phi_{p2}(\mathbf{h}_i, \mathbf{h}_j, ||\mathbf{x}_i - \mathbf{x}_j||, \mathbf{m}_{ji}, t), \tag{9}$$

$$\mathbf{x}_i \leftarrow \mathbf{x}_i + (\Delta\mathbf{x}_{K,i} + \Delta\mathbf{x}_{L,i}) \cdot \mathbb{1}_{\text{mol}}, \tag{10}$$

where $\mathbb{1}_{\text{mol}}$ is the indicator of ligand atoms since we assume the protein atoms are fixed as the context.

In practice, the condition encoder consists of two SE(3)-equivariant layers and the diffusion-based decoder consists of six SE(3)-equivaraint layers. In each SE(3)-equivariant layer, following Guan et al. (2023b), we apply graph attention to aggregate the message of each node/edge. The key/value/query embedding is obtained through a 2-layer MLP with LayerNorm and ReLU activation. Stacking these three layers as a block, our model consists of 6 blocks with `hidden_dim=128` and `n_heads=16`. Additionally, the diffusion-based decoder also have two prediction heads (which are simply 2-layer MLPs and following Softmax function) that maps the learned hidden states of atoms and edges to the predicted atom type and bond type.

### A.3 Training Details

Given a pair of protein and ligand molecule, we first decompose the molecule to get the arms. We add noise to the ligand molecules in the training set to get the perturbed molecules as the forward process of diffusion models (equation 1). The forward process is a Markov chain with fixed variance schedule $\{\beta_t\}_{t=1,\ldots,T}$ (Ho et al., 2020). We denote $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. More specifically, the noises at time $t$ are injected as follows:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}), \tag{11}$$

$$q(\mathbf{v}_t|\mathbf{v}_0) = \mathcal{C}(\mathbf{v}_t|\bar{\alpha}_t\mathbf{v}_0 + (1 - \bar{\alpha}_t)/K_a), \tag{12}$$

$$q(\mathbf{b}_t|\mathbf{b}_0) = \mathcal{C}(\mathbf{v}_t|\bar{\alpha}_t\mathbf{b}_0 + (1 - \bar{\alpha}_t)/K_b), \tag{13}$$

where $K_a$ and $K_b$ are the number of atom classes and bond classes respectively.

Then the arms and subpockets are input to the condition encoder. The output of condition encoder and the perturbed ligand are further input to the diffusion-based decoder. Then the reconstruction

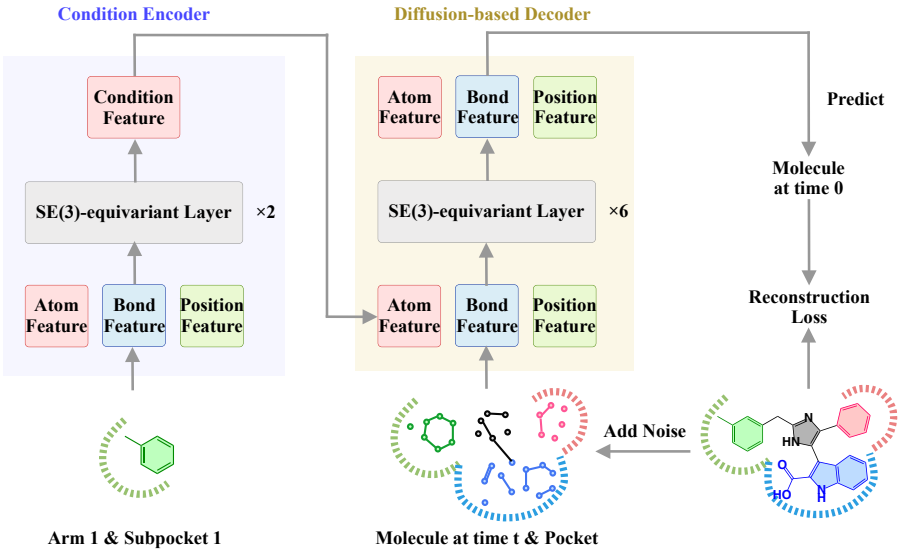


Figure 5: Illustration of training. For this case, there are actually three pairs of arms and subpockets input to the condition encoders separately. For brevity, we only plot one as an example.

loss $L_t$ at time $t$ is defined as follows:

$$L_t^{(v)} = \sum_{k=1}^{K_a} \boldsymbol{c}(\mathbf{v}_t, \mathbf{v}_0)_k \log \frac{\boldsymbol{c}(\mathbf{v}_t, \mathbf{v}_0)_k}{\boldsymbol{c}(\mathbf{v}_t, \hat{\mathbf{v}}_0)_k}, \tag{14}$$

$$L_t^{(b)} = \sum_{k=1}^{K_b} \boldsymbol{c}(\mathbf{b}_t, \mathbf{b}_0)_k \log \frac{\boldsymbol{c}(\mathbf{b}_t, \mathbf{b}_0)_k}{\boldsymbol{c}(\mathbf{b}_t, \hat{\mathbf{b}}_0)_k}, \tag{15}$$

$$L_t^{(x)} = ||\mathbf{x}_0 - \hat{\mathbf{x}}_0||^2, \tag{16}$$

$$L_t = L_t^{(x)} + \gamma_v L_t^{(v)} + \gamma_b L_t^{(b)}, \tag{17}$$

where $(\mathbf{x}_t, \mathbf{v}_t, \mathbf{b}_t)$, $(\mathbf{x}_0, \mathbf{v}_t, \mathbf{b}_t)$, and $(\hat{\mathbf{x}}_0, \hat{\mathbf{v}}_0, \hat{\mathbf{b}}_0)$ represents atom positions, atom types, and bond types of the perturbed molecule at time $t$, ground truth molecule, and the predicted molecule respectively, $\boldsymbol{c}(\mathbf{v}_t, \mathbf{v}_0) = \boldsymbol{c}^\star / \sum_{k=1}^{K_a} c_k^\star$ and $\boldsymbol{c}^\star(\mathbf{v}_t, \mathbf{v}_0) = [\alpha_t \mathbf{v}_t + (1-\alpha_t)/K_a] \odot [\bar{\alpha}_{t-1}\mathbf{v}_0 + (1-\bar{\alpha}_{t-1})/K_a]$, $\boldsymbol{c}(\mathbf{b}_t, \mathbf{b}_0) = \boldsymbol{c}^\star / \sum_{k=1}^{K_b} c_k^\star$ and $\boldsymbol{c}^\star(\mathbf{b}_t, \mathbf{b}_0) = [\alpha_t \mathbf{b}_t + (1-\alpha_t)/K_b] \odot [\bar{\alpha}_{t-1}\mathbf{b}_0 + (1-\bar{\alpha}_{t-1})/K_b]$. Note that the condition encoder and the diffusion-based decoder are jointly trained.

In practice, we set the loss weights as $\gamma_v = 100$ and $\gamma_b = 100$. Follwing the setting of Guan et al. (2023b), we set the number of diffusion steps as 1000. For this diffusion noise schedule, we choose to use a sigmoid $\beta$ schedule with $\beta_1 = 1\text{e}{-7}$ and $\beta_T = 2\text{e}{-3}$ for atom coordinates, and a cosine $\beta$ schedule suggested in Nichol & Dhariwal (2021) with $s = 0.01$ for atom types and bond types.

We use Adam Kingma & Ba (2014) with `init_learning_rate=0.0005`, `betas=(0.95, 0.999)` to train the model. And we set `batch_size=16` and `clip_gradient_norm=8`. During the training phase, we add a small Gaussian noise with a standard deviation of 0.1 to protein atom coordinates as data augmentation. We also schedule to decay the learning rate exponentially with a factor of 0.6 and a minimum learning rate of 1e-6. The learning rate is decayed if there is no improvement for the validation loss in 10 consecutive evaluations. The evaluation is performed for every 1000 training steps. We trained our model on one NVIDIA GeForce GTX A100 GPU, and it could converge within 237k steps.

### A.4 Sampling Details

To sample molecules using the pre-trained controllable and decomposed diffusion model, assume that there are available arms as conditions, we can first sample a noisy molecule from the prior

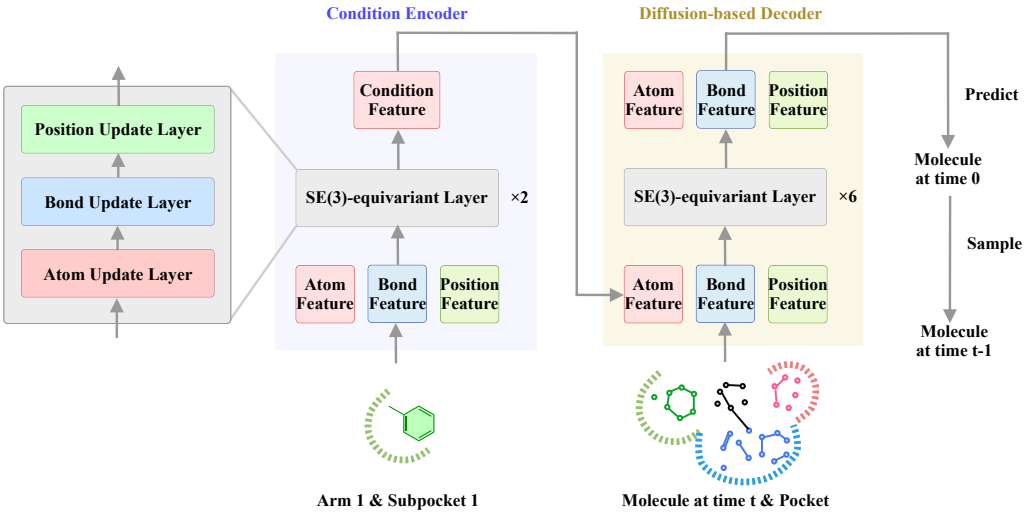


Figure 6: Illustration of the sampling.

distribution and derive a molecule by iteratively denoising following the reverse process (equation 2). More specifically, the denoising step at time $t$ corresponds to sampling molecules from the following distributions:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{x}}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \hat{\mathbf{x}}_0), \tilde{\beta}_t \mathbf{I}), \tag{18}$$

$$q(\mathbf{v}_{t-1}|\mathbf{v}_t, \hat{\mathbf{v}}_0) = \mathcal{C}(\mathbf{v}_{t-1}|\tilde{\boldsymbol{c}}_t(\mathbf{v}_t, \hat{\mathbf{v}}_0)), \tag{19}$$

$$q(\mathbf{b}_{t-1}|\mathbf{b}_t, \hat{\mathbf{b}}_0) = \mathcal{C}(\mathbf{b}_{t-1}|\tilde{\boldsymbol{c}}_t(\mathbf{b}_t, \hat{\mathbf{b}}_0)), \tag{20}$$

where $\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \hat{\mathbf{x}}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t$, $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, $\tilde{\boldsymbol{c}}(\mathbf{v}_t, \hat{\mathbf{v}}_0) = \tilde{\boldsymbol{c}}^\star / \sum_{k=1}^{K_a} \tilde{c}_k^\star$ and $\tilde{\boldsymbol{c}}^\star(\mathbf{v}_t, \hat{\mathbf{v}}_0) = [\alpha_t \mathbf{v}_t + (1-\alpha_t)/K_a] \odot [\bar{\alpha}_{t-1}\hat{\mathbf{v}}_0 + (1-\bar{\alpha}_{t-1})/K_a]$, $\tilde{\boldsymbol{c}}(\mathbf{b}_t, \hat{\mathbf{b}}_0) = \tilde{\boldsymbol{c}}^\star / \sum_{k=1}^{K_b} \tilde{c}_k^\star$ and $\tilde{\boldsymbol{c}}^\star(\mathbf{b}_t, \hat{\mathbf{b}}_0) = [\alpha_t \mathbf{b}_t + (1-\alpha_t)/K_b] \odot [\bar{\alpha}_{t-1}\hat{\mathbf{b}}_0 + (1-\bar{\alpha}_{t-1})/K_b]$. Here $(\hat{\mathbf{x}}_0, \hat{\mathbf{v}}_0, \hat{\mathbf{b}}_0)$ is the molecule output by the diffusion-based decoder, whose input is the noisy molecule at time $t$ and the condition feature. The sampling step is illustrated as Figure 6. During sampling, we also apply validity guidance proposed by Guan et al. (2023b), which encourages the model to generate molecules with valid structures.

### A.5 Optimization Details

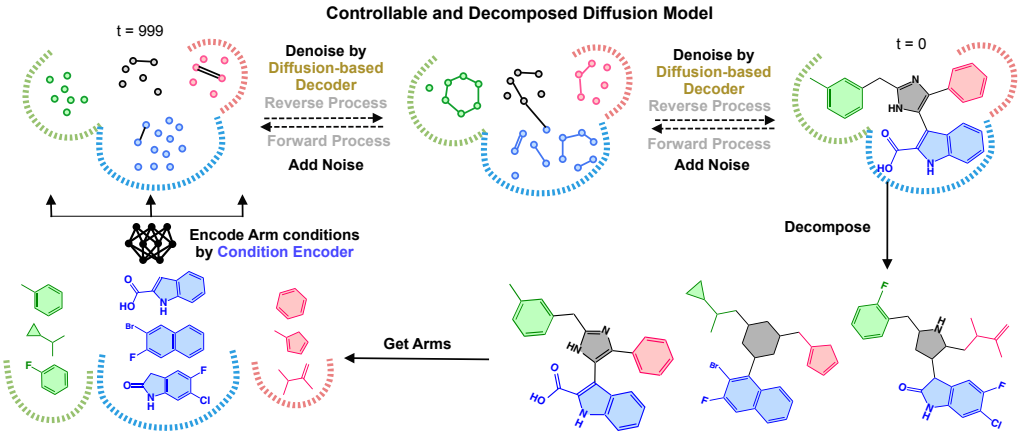


Figure 7: Illustration of molecular optimization (revised based on Figure 2). It is highlighted where we apply the condition encoder and the diffusion-based decoder.

To generate molecules with desired properties, we can apply the pre-trained controllable and decomposed diffusion models for structure-based molecular optimization without any fine-tuning. The optimization procedure is summarized as Algorithm 1 and illustrated as Figure 7. In practice, since reference ligands are not available, we can initialize the arm lists with 20 ligands generated by DecompDiff, and this is the actual setting in our experiment. We have provided the optimization procedure in detail that can be found in Section 3.2 and Section 4.1.

## B Evaluation of Molecular Conformation

To evaluate generated molecules from the perspective of molecular conformation, we compute the Jensen-Shannon divergences (JSD) in atom distance distributions between the reference molecules and the generated molecules (see Figure 8).

We also compute different bond distance and bond angle distributions of the generated molecules and compare them against the corresponding reference empirical distributions in Tables 5 and 6.

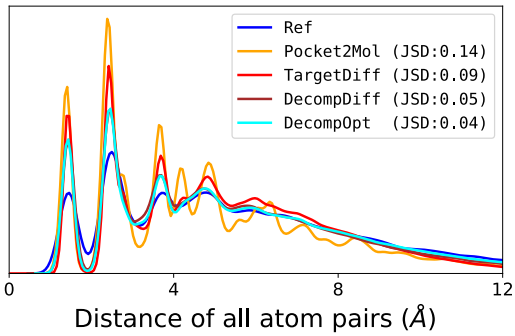


Figure 8: Comparing the distribution for distances of all-atom for reference molecules in the test set and model-generated molecules. Jensen-Shannon divergence (JSD) between two distributions is reported.

Table 5: Jensen-Shannon divergence between bond distance distributions of the reference molecules and the generated molecules, and lower values indicate better performances. "-", "=", and ":" represent single, double, and aromatic bonds, respectively. We highlight the best two results with **bold text** and underlined text, respectively.

| Bond | liGAN | GraphBP | AR | Pocket2 Mol | Target Diff | Decomp Diff | Ours |
|---|---|---|---|---|---|---|---|
| C−C | 0.601 | 0.368 | 0.609 | 0.496 | 0.369 | **0.359** | 0.362 |
| C=C | 0.665 | 0.530 | 0.620 | 0.561 | 0.505 | 0.537 | **0.504** |
| C−N | 0.634 | 0.456 | 0.474 | 0.416 | 0.363 | 0.344 | **0.328** |
| C=N | 0.749 | 0.693 | 0.635 | 0.629 | 0.550 | 0.584 | **0.566** |
| C−O | 0.656 | 0.467 | 0.492 | 0.454 | 0.421 | 0.376 | **0.373** |
| C=O | 0.661 | 0.471 | 0.558 | 0.516 | 0.461 | 0.374 | **0.329** |
| C:C | 0.497 | 0.407 | 0.451 | 0.416 | 0.263 | 0.251 | **0.196** |
| C:N | 0.638 | 0.689 | 0.552 | 0.487 | 0.235 | 0.269 | **0.219** |

To further measure the quality of generated conformation, we optimize the generated structures with Merck Molecular Force Field (MMFF) (Halgren, 1996) and calculate the energy difference between pre- and pos- MMFF-optimized coordinates for different rigid fragments that do not contain any rotatable bonds. As Table 7 and Figure 9 show, DecompOpt achieves low energy differences and outperforms baselines in most cases. We also calculate the energy difference before and after force field optimization for the whole molecules. As Table 8 and Figure 10 show, notably, DecompOpt outperforms all diffusion-based methods by a large margin and achieve comparable performance with the best baseline. These results show that the conformation of ligands generated by DecompOpt is high-quality and stable.

Table 6: Jensen-Shannon divergence between bond angle distributions of the reference molecules and the generated molecules, and lower values indicate better performances. We highlight the best two results with **bold text** and underlined text, respectively.

| Angle | liGAN | GraphBP | AR | Pocket2 Mol | Target Diff | Decomp Diff | Ours |
|---|---|---|---|---|---|---|---|
| CCC | 0.598 | 0.424 | 0.340 | 0.323 | 0.328 | 0.314 | **0.280** |
| CCO | 0.637 | 0.354 | 0.442 | 0.401 | 0.385 | 0.324 | **0.331** |
| CNC | 0.604 | 0.469 | 0.419 | **0.237** | 0.367 | 0.297 | 0.280 |
| OPO | 0.512 | 0.684 | 0.367 | 0.274 | 0.303 | 0.217 | **0.198** |
| NCC | 0.621 | 0.372 | 0.392 | 0.351 | 0.354 | 0.294 | **0.266** |
| CC=O | 0.636 | 0.377 | 0.476 | 0.353 | 0.356 | 0.259 | **0.257** |
| COC | 0.606 | 0.482 | 0.459 | **0.317** | 0.389 | 0.339 | 0.338 |

Table 7: Median energy difference for rigid fragment of different fragment size (3/4/5/6/7/8 atoms) before and after the force-field optimization.

| Methods | Median Energy Difference (↓) | | | | | |
|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 |
| LiGAN | 86.32 | 165.15 | 105.96 | 185.70 | 243.79 | 332.81 |
| AR | 25.79 | 73.06 | 23.89 | 30.42 | 56.47 | 76.50 |
| Pocket2Mol | 10.43 | 33.93 | 34.47 | 27.86 | 33.90 | 42.97 |
| TargetDiff | 7.31 | 30.57 | 18.01 | 11.98 | 28.92 | 50.42 |
| DecompDiff | 6.01 | 29.20 | 10.78 | 4.33 | 12.74 | 30.68 |
| DECOMPOPT | 6.00 | 16.59 | 9.89 | 2.61 | 13.29 | 31.49 |

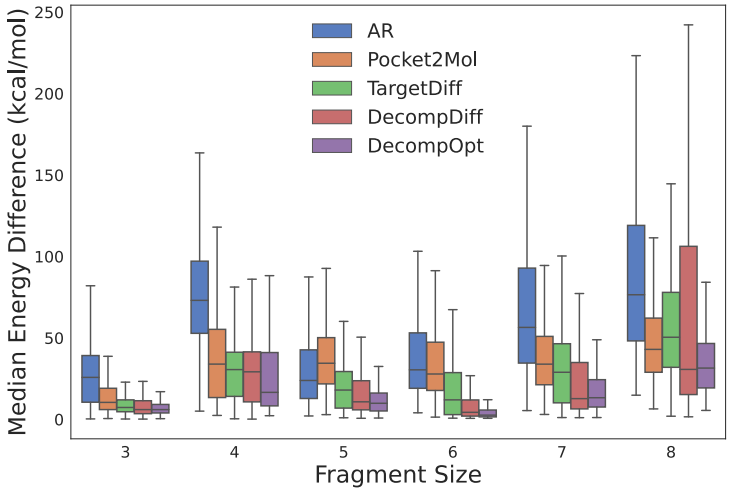


Figure 9: Median energy difference for molecules with different number of rotatable bonds before and after the force-field optimization.

## C ADDITIONAL RESULTS

### C.1 FULL EVALUATION RESULTS

Table 8: Median energy difference for molecules with different number of rotatable bonds (1/2/3/4/5/6/7 rotatable bonds) before and after the force-field optimization.

| Methods | Median Energy Difference (↓) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| LiGAN | 810.45 | 981.53 | 1145.96 | 1783.95 | 1960.24 | 2547.32 | 2735.75 |
| AR | 176.67 | 222.74 | 244.51 | 268.01 | 332.89 | 388.70 | 441.90 |
| Pocket2Mol | 105.64 | 125.19 | 168.84 | 199.33 | 204.82 | 226.73 | 263.96 |
| TargetDiff | 225.48 | 253.72 | 303.60 | 344.12 | 360.74 | 420.47 | 434.30 |
| DecompDiff | 279.44 | 264.16 | 268.23 | 265.57 | 262.69 | 279.73 | 289.07 |
| DECOMPOPT | 63.33 | 169.17 | 215.19 | 248.35 | 202.81 | 237.38 | 238.32 |

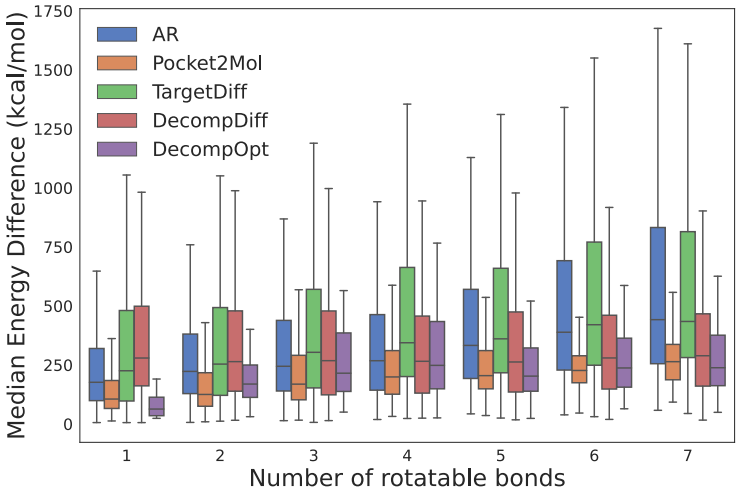


Figure 10: Median energy difference for molecules with different number of rotatable bonds before and after the force-field optimization.

We provide box plots of evaluation metrics as shown in Figure 11.

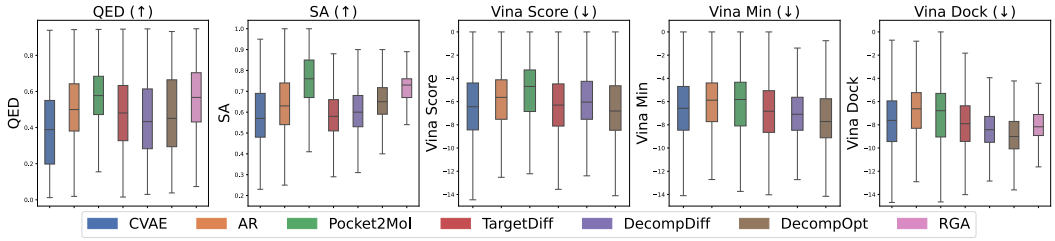


Figure 11: The boxplots of QED, SA, Vina Score, Vina Minimize, and Vina Dock of ligands generated by DECOMPOPT and baseline models.

Following Guan et al. (2023b), our model also has variants of priors. Table 9 shows the results of multiple variants of our models. The setting of *Ref Prior*, *Pocket Prior*, and *Opt Prior* strictly follows DecompDiff (Guan et al., 2023b). *Ref Best* means using the best checkpoint instead of the last checkpoint for each target pocket during optimization with reference priors for evaluation. For *Pocket/Opt Best*, it is similar. *Best of Best* means using the best checkpoint across all checkpoints with *Ref Prior* and *Pocket Prior* during optimization for each target pocket.

Table 9: Summary of different properties of reference molecules and molecules generated by our model and other generation (Gen.) and optimization (Opt.) baselines. (↑) / (↓) denotes a larger / smaller number is better.

| Methods | Vina Score (↓) | | Vina Min (↓) | | Vina Dock (↓) | | High Affinity (↑) | | QED (↑) | | SA (↑) | | Diversity (↑) | | Success |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Rate |
| Reference | -6.36 | -6.46 | -6.71 | -6.49 | -7.45 | -7.26 | - | - | 0.48 | 0.47 | 0.73 | 0.74 | - | - | 25.0% |
| DECOMPOPT (Ref Prior) | -5.68 | -5.88 | -6.53 | -6.49 | -7.49 | -7.66 | 59.2% | 65.0% | 0.56 | 0.58 | 0.73 | 0.73 | 0.64 | 0.66 | 35.4% |
| DECOMPOPT (Ref Best) | -5.75 | -5.97 | -6.58 | -6.70 | -7.63 | -8.02 | 62.6% | 74.3% | 0.56 | 0.59 | 0.73 | 0.72 | 0.63 | 0.67 | 39.4% |
| DECOMPOPT (Pocket Prior) | -5.27 | -6.38 | -7.07 | -7.45 | -8.85 | -8.72 | 71.4% | 93.8% | 0.40 | 0.36 | 0.63 | 0.63 | 0.60 | 0.61 | 29.2% |
| DECOMPOPT (Pocket Best) | -5.33 | -6.49 | -7.08 | -7.60 | -9.01 | -8.98 | 73.9% | 100% | 0.41 | 0.39 | 0.63 | 0.63 | 0.59 | 0.60 | 44.7% |
| DECOMPOPT (Opt Prior) | -5.73 | -6.64 | -7.29 | -7.53 | -8.78 | -8.72 | 70.3% | 89.9% | 0.46 | 0.44 | 0.65 | 0.65 | 0.61 | 0.61 | 38.1% |
| DECOMPOPT (Opt Best) | -5.87 | -6.81 | -7.35 | -7.72 | -8.98 | -9.01 | 73.5% | 93.3% | 0.48 | 0.45 | 0.65 | 0.65 | 0.60 | 0.61 | 52.5% |
| DECOMPOPT (Best of Best) | -6.22 | -6.94 | -7.50 | -7.74 | -8.98 | -8.95 | 76.2% | 100% | 0.51 | 0.51 | 0.67 | 0.67 | 0.61 | 0.63 | 60.6% |

### C.2 TRADE-OFF BETWEEN SUCCESS RATE AND DIVERSITY

In addition to overall performance, we also show the trade-off between Success Rate and diversity of RGA, TargetDiff w/ Opt., and DECOMPOPT for each target protein pocket in Figure 12. DECOMPOPT shows general superiority to the other two baselines in most cases considering both Success Rate and diversity.

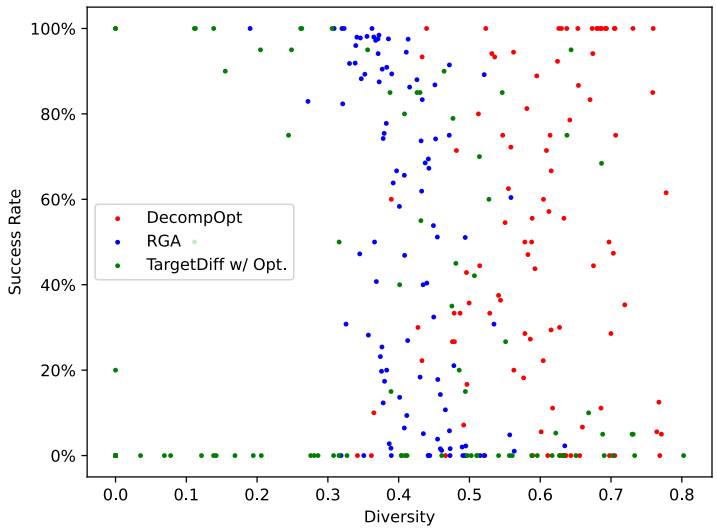


Figure 12: Trade-off of Success Rate and diversity. Each point with coordinate $(x, y)$ represents a pocket with Success Rate $x$ and diversity $y$. The closer to the top right, the better.

### C.3 EVALUATION OF THE ABILITY TO DESIGN NOVEL LIGANDS

We additionally test the **Novelty** and **Similarity** of generated ligands compared with the reference ligand. Novelty is defined as the ratio of generated ligands that are different from the reference ligands of the corresponding pockets in the test set. Similarity is defined as the Tanimoto Similarity between the generated ligands and the corresponding reference ligands. The results show that the generated ligands are not similar to reference ligands in the test set. Besides, we also test **Uniqueness** and **Diversity** of generated ligands. Uniqueness is the percentage of unique molecules among all the generated molecules. Diversity is the same as that in Section 4.1. The results are reported in Table 10. These results show that DECOMPOPT can design novel ligands, which is an important ability for drug discovery.

Table 10: Evaluation of the ability to design novel ligands.

| Methods | Novelty | Similarity | Uniqueness | Diversity |
|---|---|---|---|---|
| LiGAN | 100% | 0.22 | 87.82% | 0.66 |
| AR | 100% | 0.24 | 100% | 0.70 |
| Pocket2Mol | 100% | 0.26 | 100% | 0.69 |
| TargetDiff | 100% | 0.30 | 99.63% | 0.72 |
| DecompDiff | 100% | 0.34 | 99.99% | 0.68 |
| RGA | 100% | 0.37 | 96.82% | 0.41 |
| DECOMPOPT | 100% | 0.36 | 100% | 0.60 |

### C.4 EFFECTS OF SUBPOCKETS.

To study the influence of subpockets in controlling the optimization, we further conducted an ablation study using only arms without subpockets as conditions. As Table 11 shows, while DECOMPOPT, when solely with arms as conditions, is capable of optimizing all metrics, its efficiency in this scenario is not as well as DECOMPOPT that utilizes both arms and pockets as conditions. Recall that we use SE(3)-invaraint features of arms (and subpockets) as conditions. Without subpockets, this feature would be agnostic to the molecular interaction and spatial relation between the arms and subpockets. Such information is important to some of the properties (e.g., Vina scores). The SE(3)-invaraint features from pairs of subpockets and arms contain the aforementioned information and are better aligned with the protein-ligand complex being generated.

Table 11: Comparison of DECOMPOPT optimization results with only arms and arm-pocket complexes as conditions. (↑) / (↓) denotes a larger / smaller number is better.

| Methods | Vina Score (↓) | | Vina Min (↓) | | Vina Dock (↓) | | High Affinity (↑) | | QED (↑) | | SA (↑) | | Diversity (↑) | | Success |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Rate |
| DecompDiff | -5.67 | -6.04 | -7.04 | -7.09 | -8.39 | -8.43 | 64.4% | 71.0% | 0.45 | 0.43 | 0.61 | 0.60 | 0.68 | 0.68 | 24.5% |
| DECOMPOPT (arms-only) | -5.52 | -6.26 | -7.05 | -7.26 | -8.65 | -8.64 | 66.6% | 86.1% | 0.46 | 0.43 | 0.63 | 0.63 | 0.63 | 0.63 | 45.7% |
| DECOMPOPT | -5.87 | -6.81 | -7.35 | -7.72 | -8.98 | -9.01 | 73.5% | 93.3% | 0.48 | 0.45 | 0.65 | 0.65 | 0.60 | 0.61 | 52.5% |

### C.5 INFLUENCE OF THE QUALITY OF INITIAL LIGANDS ON PERFORMANCE.

To study the influence of the quality of initial ligands on performance of structure-based molecular optimization, we have conducted an ablation study focusing on Vina Min score optimization, using ligands with high and low Vina Min scores as initializations for the arm lists. Due to limited resources, we chose to conduct this study on the protein 2V3R, which is randomly chosen from our test set. We generated 100 ligands using DecompDiff and selected 20 ligands with the highest and lowest Vina Min scores. These ligands were then used as the initial conditions for the optimization process. As shown in Table 12, the optimization outcomes are slightly influenced by the quality of the initial ligands. However, regardless the quality of the initial ligands, DECOMPOPT can consistently improve the quality of the generated ligands.

Table 12: Comparison of optimization with initial ligands of different quality.

| | High Vina Min Scores | | Low Vina Min Scores | | Δ (high - low) |
|---|---|---|---|---|---|
| | Avg. | Med. | Avg. | Med. | Avg. |
| Initial ligands | -8.54 | -8.48 | -7.08 | -7.04 | -1.46 |
| DECOMPOPT | -9.12 | -8.96 | -9.00 | -8.96 | -0.12 |

### C.6 INFLUENCE OF THE NUMBER OF INITIAL LIGANDS ON PERFORMANCE.

To study the influence of the number of initial molecules on the performance of structure-based molecular optimization, we further run the experiments with initial arm lists of 1 and 5 molecules generated by DecompDiff. As Table 13 indicates, the initial number of molecules has a modest impact on the optimization outcomes, with a higher number of molecules generally leading to improved performance. Notably, even when starting with a single molecule generated by DecompDiff, DECOMPOPT demonstrates a considerably high success rate.

Table 13: Summary of results using different number of molecules to initialize arm lists. (↑) / (↓) denotes a larger / smaller number is better.

| Methods | Vina Score (↓) | | Vina Min (↓) | | Vina Dock (↓) | | High Affinity (↑) | | QED (↑) | | SA (↑) | | Diversity (↑) | | Success |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Rate |
| init num = 1 | -5.41 | -6.61 | -7.12 | -7.51 | -8.78 | -8.82 | 70.4% | 88.9% | 0.47 | 0.45 | 0.64 | 0.63 | 0.61 | 0.61 | 47.0% |
| init num = 5 | -5.71 | -6.71 | -7.25 | -7.58 | -8.86 | -8.97 | 71.8% | 93.3% | 0.49 | 0.46 | 0.65 | 0.64 | 0.60 | 0.61 | 49.4% |
| init num = 20 | -5.87 | -6.81 | -7.35 | -7.72 | -8.98 | -9.01 | 73.5% | 93.3% | 0.48 | 0.45 | 0.65 | 0.65 | 0.60 | 0.61 | 52.5% |

# D EXTENDED RESULTS OF CONTROLLABILITY

## D.1 R-GROUP OPTIMIZATION

We provide additional R-group Optimization experiment on protein 4G3D, as shown in Figure 13.

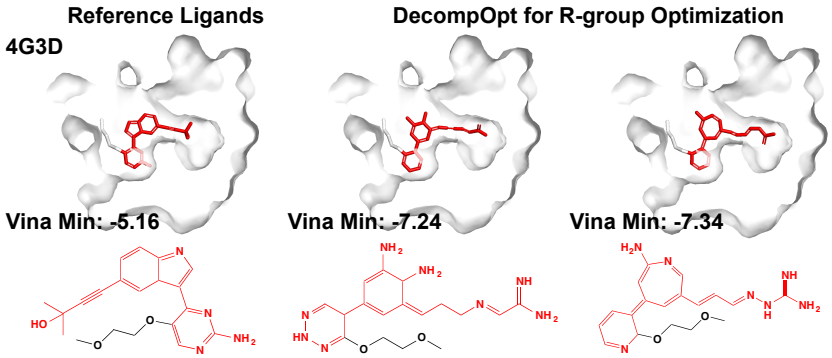


Figure 13: Additional R-group optimization result. The left column is reference binding molecule, the middle and right columns are molecules generated by DECOMPOPT with 30 rounds of optimization on protein 4G3D. Optimized R-group are highlighted in red.

Table 14: R-group optimization results generated using Decompdiff and DECOMPOPT on protein 3DAF and 4F1M. DECOMPOPT was optimized over 30 rounds towards high Vina Min Score and evaluated using the final round results. Both targets were assessed with 20 generated molecules and the mean of properties are reported.

| Model | 3DAF | | | 4F1M | | |
|---|---|---|---|---|---|---|
| | Vian Min (↓) | Tanimoto Sim. (↑) | Complete. (↑) | Vian Min (↓) | Tanimoto Sim. (↑) | Complete. (↑) |
| Decompdiff | -8.44 | 0.15 | 60.0% | -5.90 | 0.15 | 65.0% |
| DECOMPOPT | -9.39 | 0.23 | 95.0% | -6.32 | 0.49 | 55.0% |

## D.2 FRAGMENT GROWING

Enhancing the binding affinity of drug candidates through combination of R-group optimization and fragment growing can effectively leverage capabilities of DECOMPOPT. The quantitative results are shown in Table 14. For our case study, we perform R-group optimization and fragment growing on 5AEH. Starting from a high binding affinity drug candidate, we first optimize R-group for 30 rounds same as workflow in Section 4. Subsequently, we design the new arms prior and atom num with expert guidance, and expand fragments using DECOMPOPT. As Figure 14 shows, DECOMPOPT ultimately generates molecules with a Vina Min Score more than 4 kcal/mol better than the reference.

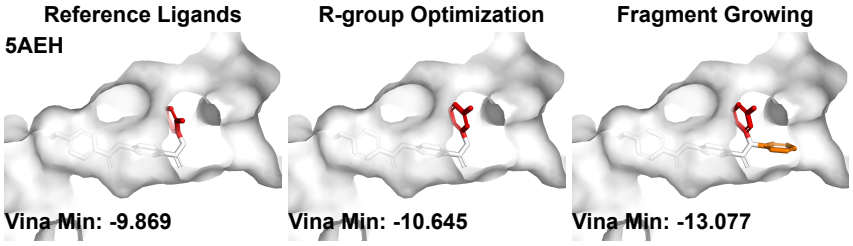


Figure 14: Example of R-group optimization and fragment growing conducted using DECOMPOPT on 5AEH. The reference ligand, the best R-group result, and the best fragment growing result based on R-group optimization are displayed from left to right. The selected R-group is highlighted in red, while the newly extended arm is highlighted in orange.

### D.3 SCAFFOLD HOPPING

**Additional Evaluation Metrics** In addition to evaluation metrics discussed in Section 4, we evaluated *Validity*, *Uniqueness*, *Novelty*, *Complete Rate*, and *Scaffold Similarity* to measure models' capability in scaffold hopping. Detailed calculation of these metrics as follows:

- **Validity** is defined as the fraction of generated molecules that can be successfully sanitized.
- **Uniqueness** measures the proportion of unique molecules among the generated molecules.
- **Novelty** measures the fraction of generated molecules that not presented in training set.
- **Complete Rate** measures the proportion of completed molecules within the generated results.
- **Scaffold Similarity** Following Polykovskiy et al. (2020), Bemis–Murcko scaffolds are extracted using rdkit function `MurckoScaffold`. We count the occurrences of scaffolds in all generated and reference molecules, creating vectors $G$ and $R$, where each dimension represents the count of a specific scaffold. The scaffold similarity is calculated as the cosine similarity between vectors $G$ and $R$.

**More Examples of Generated Results** For scaffold hopping, we provide more visualization of ligands generated by DECOMPOPT and DecompDiff on protein 2Z3H, 4AVW, 4QLK, and 4BEL, which are shown in Figure 15.

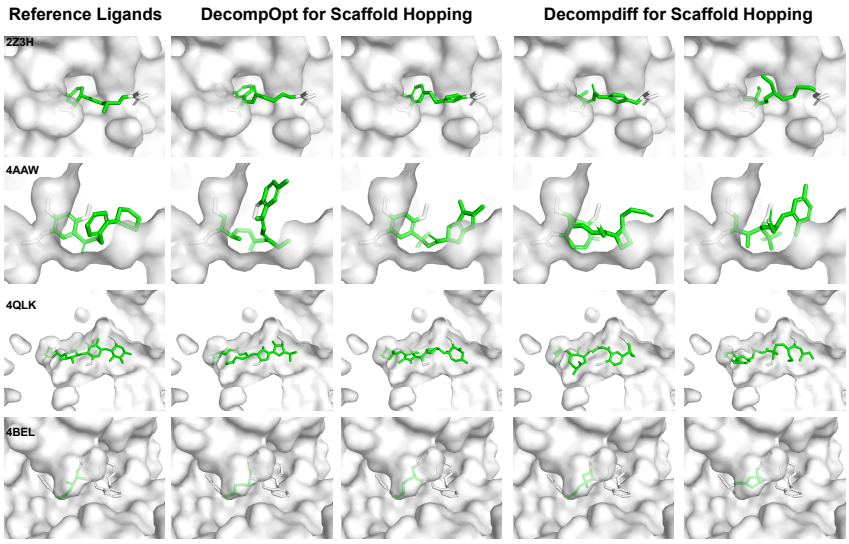


Figure 15: More examples of Scaffold Hopping results. The left column shows reference ligands, Scaffold Hopping results generated by DECOMPOPT are shown at the second and the third rows, and results generated by DecompDiff are shown at the fourth and the fifth rows. Scaffold are highlighted in green.

**Changes in Molecular Properties After Scaffold Hopping** Scaffold hopping aims at finding scaffold structures that can connect existing functional groups without disrupting their interactions with the target protein. The main purpose of this is to find novel scaffolds which are not protected by existing patents while maintaining comparable properties as the original molecule. Therefore, we did not implement property optimization mechanisms in scaffold hopping tasks and solely focusing on designing scaffolds that can connect existing arms. We provide the property comparison before and after scaffold hopping in Table 15. As the result shows, the properties of the ligands remain relatively consistent before and after the process of scaffold hopping.

Table 15: Summary of properties of reference molecules and molecules generated through scaffold hopping using DECOMPOPT. (↑) / (↓) denotes a larger / smaller number is better.

| Methods | Vina Score (↓) | | Vina Min (↓) | | Vina Dock (↓) | | QED (↑) | | SA (↑) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. | Avg. | Med. |
| Reference | -6.36 | -6.46 | -6.71 | -6.49 | -7.45 | -7.26 | 0.48 | 0.47 | 0.73 | 0.74 |
| Scaffold Hopping by DECOMPOPT | -5.89 | -6.13 | -6.46 | -6.28 | -7.28 | -7.48 | 0.49 | 0.48 | 0.71 | 0.69 |