

479 **A Proof of Proposition 1.**

480 First, we here denote all atom's positions in the molecules as $\mathbf{X}_M \in \mathbb{R}^{N_a \times 3}$, and in the protein as
 481 $\mathbf{X}_M \in \mathbb{R}^{N_{aa} \times 3}$, and linker and functional group type as $\mathbf{S}_M \in \mathbb{R}^{(N_{aa} + N_{fg}) \times (22 + M_{fg} + M_{at})}$. Note that
 482 one functional group may contain several atoms so that $N_{aa} + N_{fg} < N_a$.

483 SE(3) group as a roto-translation group in \mathbb{R}^3 , can be divided into two groups: SO(3) as the rotation
 484 group and T(3) as the translation group. For $\mathbf{x} \in \mathbb{R}^3$, and $g = r + v$ with $g \in \text{SE}(3)$, $r \in \text{SO}(3)$, $v \in$
 485 $\text{T}(3)$, $T_g(\mathbf{x}) = T_{r+v}(\mathbf{x}) = T_v \circ T_r(\mathbf{x})$

486 **Lemma A1.** If the The equivariance and invariance of the distribution in the reverse diffusion process
 487 are listed as

$$\begin{aligned} p(T_g(\mathbf{X}_M^T), \mathbf{S}_M^T | T_g(\mathbf{X}_P), \mathbf{S}_P) &= p(\mathbf{X}_M^T, \mathbf{S}_M^T) = p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P) \\ p(T_g(\mathbf{X}_M^{t-1}) | T_g(\mathbf{X}_M^t), \mathbf{S}_M^t, T_g(\mathbf{X}_P), \mathbf{S}_P) &= p(\mathbf{X}_M^{t-1} | \mathbf{X}_M^t, \mathbf{S}_M^t, \mathbf{X}_P, \mathbf{S}_P) \\ p(\mathbf{S}_M^{t-1} | T_g(\mathbf{X}_M^t), \mathbf{S}_M^t, T_g(\mathbf{X}_P), \mathbf{S}_P) &= p(\mathbf{S}_M^{t-1} | \mathbf{X}_M^t, \mathbf{S}_M^t, \mathbf{X}_P, \mathbf{S}_P), \end{aligned} \quad (13)$$

488 Then $p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P)$ is SE(3) invariant.

489 *Proof.* Since $\mathbf{T}_g(\mathcal{M}) = \{\mathcal{S}, T_g(\mathbf{X}_M)\}$, we can write the joint generative process as

$$\begin{aligned} &p(T_g(\mathbf{X}_M), \mathbf{S}_M | T_g(\mathbf{X}_P), \mathbf{S}_P) \\ &= \int p(T_g(\mathbf{X}_M^T), \mathbf{S}_M^T | T_g(\mathbf{X}_P), \mathbf{S}_P) \prod_{t=0}^{T-1} p(T_g(\mathbf{X}_M^{t-1}), \mathbf{S}_M^{t-1} | T_g(\mathbf{X}_M^t), \mathbf{S}_M^t, T_g(\mathbf{X}_P), \mathbf{S}_P) d\mathcal{M}^{0:T-1} \\ &= \int p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P) \prod_{t=0}^{T-1} p(T_g(\mathbf{X}_M^{t-1}), \mathbf{S}_M^{t-1} | T_g(\mathbf{X}_M^t), \mathbf{S}_M^t, T_g(\mathbf{X}_P), \mathbf{S}_P) d\mathcal{M}^{0:T-1} \\ &= \int p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P) \prod_{t=0}^{T-1} p(T_g(\mathbf{X}_M^{t-1}) | T_g(\mathcal{M}^t, \mathcal{P})) p(\mathbf{S}_M^{t-1} | T_g(\mathcal{M}^t, \mathcal{P})) d\mathcal{M}^{0:T-1} \\ &= \int p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P) \prod_{t=0}^{T-1} p(\mathbf{X}_M^{t-1} | \mathcal{M}^t, \mathcal{P}) p(\mathbf{S}_M^{t-1} | \mathcal{M}^t, \mathcal{P}) d\mathcal{M}^{0:T-1} \\ &= \int p(\mathbf{X}_M^T, \mathbf{S}_M^T | \mathbf{X}_P, \mathbf{S}_P) \prod_{t=0}^{T-1} p(\mathbf{X}_M^{t-1}, \mathbf{S}_M^{t-1} | \mathbf{X}_M^t, \mathbf{S}_M^t, \mathbf{X}_P, \mathbf{S}_P) d\mathcal{M}^{0:T-1} \\ &= p(\mathbf{X}_M, \mathbf{S}_M | \mathbf{X}_P, \mathbf{S}_P) \end{aligned} \quad (14)$$

490 Then, let's consider a single atom's position. We here denote each atom's \mathbf{x}_M^t as

$$\mathbf{x}_M^t = \mathbf{x}_C^t + \mathbf{x}_R^t \mathbf{O}_C^t \quad (15)$$

491 where \mathbf{x}_C^t is the defined center atom's position in the functional group, \mathbf{x}_R^t is the relative position
 492 of the atom in the local coordinate system centering at \mathbf{x}_C^t , \mathbf{O}_C^t is the rotation matrices of the local
 493 coordinate system with respect to the global system. Moreover, because the functional group is
 494 regarded as rigid bodies, $\mathbf{x}_R^t = \mathbf{x}_R$ is constant. To be specific, if \mathbf{x}_M^t refers to linker's position,
 495 $\mathbf{O}_C^t = \mathbf{0}$.

496 **Proposition A2.** If each atom's relative positions in the local coordinate systems are fixed, and
 497 $p(\mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P})$ is SE(3)-equivariant and $p(\mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P})$ is SO(3)-equivariant and T(3)-invariant,
 498 such that $p(T_g(\mathbf{x}_C^{t-1}) | T_g(\mathcal{M}^t, \mathcal{P})) = p(\mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P})$ and $p(T_r(\mathbf{O}_C^{t-1}) | T_g(\mathcal{M}^t, \mathcal{P})) =$
 499 $p(\mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P})$, where $r \in \text{SO}(3)$, $v \in \text{T}(3)$, $r + v = g \in \text{SE}(3)$, then $p(\mathbf{x}_M^{t-1} | \mathcal{M}^t, \mathcal{P})$ is
 500 SO(3)-equivariant.

501 *Proof.* According to the convolution formula in probability theory, if $w = u + v$, then

$$p(w) = \int p(u, w - u) du = \int p(w - v, v) dv \quad (16)$$

502 By using the Eq. (16), we can write every single atom’s position density function as

$$\begin{aligned}
& p(\mathbf{x}_M^{t-1} | \mathcal{M}^t, \mathcal{P}) \\
& = p(\mathbf{x}_C^{t-1} + \mathbf{x}_R \mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) \\
& = \int p(\mathbf{x}_C^{t-1}, \mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) d\mathbf{x}_C^{t-1} \\
& = \int p(\mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) p(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) d\mathbf{x}_C^{t-1}
\end{aligned} \tag{17}$$

503 Since

$$p(\mathbb{T}_g(\mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) = p(\mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}), \tag{18}$$

504 and

$$p(\mathbb{T}_r(\mathbf{O}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) = p(\mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P}). \tag{19}$$

505 We can obtain that

$$\begin{aligned}
& p(\mathbb{T}_r(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) \\
& = p(\mathbb{T}_r(\mathbf{x}_R \mathbf{O}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) \\
& = \frac{1}{\mathbf{x}_R} p(\mathbb{T}_r(\mathbf{O}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) \\
& = \frac{1}{\mathbf{x}_R} p(\mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) \\
& = p(\mathbf{x}_R \mathbf{O}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) \\
& = p(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P})
\end{aligned} \tag{20}$$

506 Therefore, according to Eq. (18), (20), and (17)

$$\begin{aligned}
& p(\mathbb{T}_g(\mathbf{x}_M^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) \\
& = \int p(\mathbb{T}_g(\mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) p(\mathbb{T}_{r+v}(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) d\mathbf{x}_C^{t-1} \\
& = \int p(\mathbb{T}_g(\mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) p(\mathbb{T}_r(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1}) | \mathbb{T}_g(\mathcal{M}^t, \mathcal{P})) d\mathbf{x}_C^{t-1} \\
& = \int p(\mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) p(\mathbf{x}_M^{t-1} - \mathbf{x}_C^{t-1} | \mathcal{M}^t, \mathcal{P}) d\mathbf{x}_C^{t-1} \\
& = p(\mathbf{x}_M^{t-1} | \mathcal{M}^t, \mathcal{P})
\end{aligned} \tag{21}$$

507 *Proof of Proposition 1.* The sufficiency of SE(3)-invariance of $p(\mathbf{s}^{t-1} | \mathcal{M}^t, \mathcal{P})$ and $p(\mathbf{s}^T)$ is given in
508 **Lemma A.1**, and the sufficiency of SE(3)-equivariance of $p(\mathbf{x}^{t-1} | \mathcal{M}^t, \mathcal{P})$, and SO(3)-equivariance
509 and T(3)-invariance of $p(\mathbf{O}^{t-1} | \mathcal{M}^t, \mathcal{P})$ is given in **Proposition A.2**. Besides, it is easy to obtain that
510 if $p(\mathbf{O}^T)$ and $p(\mathbf{x}^T)$ is SE(3)-invariant distribution, then $p(\mathbf{x}_M^T)$ will be invariant.

511 B Method Details

512 **Amino acid context encoding.** Several geometric or type features are embedded to encode amino
513 acids. For the geometric features including torsion angles/dihedrals, and pairwise distances, they
514 are all roto-translational invariant, since the geometric features are all scalars obtained from relative
515 coordinates. Besides, the local coordinates of atoms in a single amino acid are also invariant because
516 it is always fixed in the local frame established by C_α , C and N. For the type features including amino
517 acid types, sequential relationships, and pair of amino acid types, the translational and rotational
518 operation is unrelated to them. Thus, the encoded amino acid contexts are roto-translational invariant,
519 leading to the invariance of all the follow-up embeddings.

520 **Equivariant neural network for linkers.** For the roto-translational equivariance of positions for
521 single atoms, since $\mathbf{O}_j^t = \mathbf{I}$, Eq. 12 will be written as $G(\mathcal{M}^t, \mathcal{P})[j] = \text{MLP}_C(\mathbf{h}'_j)$, unable to satisfy

522 the equivariance. In this way, we revised it for single atoms by using the EGNN[33] stacked in the
523 final layer for updating the positions, which reads

$$G(\mathcal{M}^t, \mathcal{P})[j] = \mathbf{x}_j + \frac{1}{C_j} \sum_{i \in \mathcal{I}_{\text{at}} \cup \mathcal{I}_{\text{fg}}} (\mathbf{x}_j - \mathbf{x}_i) \mathbf{h}'_i, \quad (22)$$

524 where we choose $C_j = \|\mathbf{x}_j\|_2 + 1$.

525 C Data Preprocessing

526 **Local frame establishment.** In 3D Euclidian space, for a rigid body including more than mass
527 points that are not co-linear, a local frame can be established. We first choose a center node (center
528 point) A as the origin, and a second node B , leading to \vec{AB} as the direction of x-axis. Then, a third
529 node C is selected. By Schmidt orthogonalization of \vec{AC} with respect to \vec{AB} , the direction of y-axis
530 can be computed. And finally, the direction of z-axis is obtained by the cross-product of the unit
531 vectors in the x direction and y direction. By this means, the local frame is established by the three
532 nodes, and the other nodes' local coordinates can be determined in the local frame. Because the local
533 frame requires at least three points to establish, the functional groups including only 2 atoms are
534 divided into two linkers.

535 **Functional group datasets.** We give detailed information on the included functional groups,
536 including 2D graph, 3D structure, time of occurrence in the CrossDocked2020, and the center node
537 (node A), node B , and node C of the functional group in Table. 8.

538 Note that in beneze, the symmetric structures lead the frame nodes to be any three consecutive
539 points on the hexagon. Besides, for the functional groups of 'NS(=O)=O' and 'O=CNO', two stable
540 conformations exist, so we in practice regard them as four different types.

541 D Experiment Details

542 **Platform.** We use a single NVIDIA A100(81920MiB) GPU for a trial. The codes are implemented
543 in Python 3.9 mainly with Pytorch 1.12, and run on Ubuntu Linux.

544 **Model Details.** In the diffusion of orientation and position, we employ a cosine variance schedule
545 for $\bar{\alpha}_t$, which reads

$$\bar{\alpha}_t = \cos^2 \left(\frac{\pi}{2} \left(\frac{t}{T} + s \right) / (1 + s) \right) / \cos^2 \left(\frac{\pi}{2} s / (1 + s) \right), \quad (23)$$

546 where $s = 0.01$. In the diffusion of atom type, β_t is set as $\beta_t = \frac{t}{T}$. For the denoiser, the layer number
547 is set as 6, and the embedding size is set as 256. The model is trained with Adam optimizer in 5000
548 epochs.

549 **Functional group and Atom type analysis.** We give a detailed analysis of functional groups and
550 atom types, in Table. 5, 6 and 7.

Table 5: Frequency of the top ten functional groups that occur most frequently in Crossdocked2020.

Functional Group	Ref.	Pocket2Mol	TargetDiff	DiffSBDD	D3FG(Joint)	D3FG(Stage)
c1ccccc1	0.392	0.491	0.277	0.007	0.372	0.409
NC=O	0.147	0.075	0.142	0.201	0.082	0.107
O=CO	0.119	0.169	0.303	0.579	0.154	0.085
c1ccncc1	0.045	0.072	0.049	0.018	0.027	0.052
c1ncc2nc[nH]c2n1	0.034	0.001	0.000	0.000	0.001	0.020
NS(=O)=O	0.030	0.000	0.000	0.008	0.001	0.001
O=P(O)(O)O	0.022	0.003	0.193	0.000	0.007	0.010
OCO	0.019	0.024	0.091	0.016	0.045	0.050
c1cncnc1	0.017	0.008	0.138	0.000	0.002	0.009
c1cn[nH]c1	0.016	0.011	0.001	0.000	0.003	0.004
JSD	-	0.248	0.301	0.553	0.223	0.201

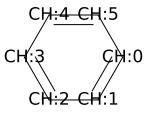
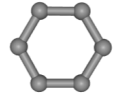
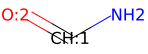
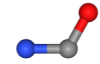
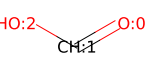
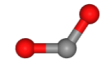
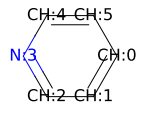
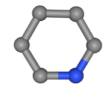
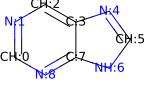
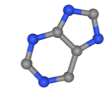



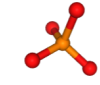
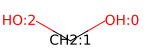
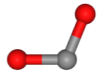

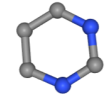
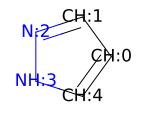
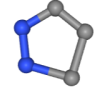
Table 6: Ratio of the atoms.

Atom	Ref.	Pocket2Mol	TargetDiff	DiffSBDD	D3FG(Joint)	D3FG(Stage)
C	15.866	14.956	17.744	13.526	17.766	15.999
N	2.765	1.956	2.192	2.236	2.157	1.943
O	4.006	2.538	4.389	3.071	3.732	3.353
F	0.309	0.084	0.239	0.160	0.193	0.170
P	0.263	0.024	0.119	0.034	0.969	0.088
S	0.266	0.038	0.104	0.149	0.169	0.153
Cl	0.152	0.016	0.064	0.006	0.145	0.122
MAE	-	0.573	0.471	0.627	0.528	0.294

Table 7: Frequency of the atoms.

Atom	Ref.	Pocket2Mol	TargetDiff	DiffSBDD	D3FG(Joint)	D3FG(Stage)
C	0.672	0.762	0.714	0.702	0.741	0.733
N	0.117	0.100	0.088	0.116	0.124	0.089
O	0.170	0.129	0.176	0.159	0.175	0.154
F	0.013	0.004	0.009	0.008	0.009	0.008
P	0.011	0.001	0.005	0.002	0.002	0.004
S	0.011	0.002	0.004	0.007	0.001	0.007
Cl	0.006	0.001	0.002	0.003	0.001	0.006
JSD	-	0.098	0.059	0.054	0.075	0.056

Table 8: The included functional groups in D3FG. ‘T’ is the occurrence times of the functional group in the datasets (100,000 ligands).

Smiles	2D graph	3D structures	A	B	C	T
<chem>c1ccccc1</chem>			1	0	2	131148
<chem>NC=O</chem>			1	0	2	49023
<chem>O=CO</chem>			1	0	2	39863
<chem>c1ccncc1</chem>			3	2	4	15115
<chem>c1ncc2nc[nH]c2n1</chem>			7	3	6	11369
<chem>NS(=O)=O</chem>			1	0	2	10121
<chem>O=P(O)(O)O</chem>			1	0	2	7451
<chem>OCO</chem>			1	0	2	6405
<chem>c1cncnc1</chem>			3	2	4	5965
<chem>c1cn[nH]c1</chem>			2	3	1	5404

Smiles	2D graph	3D structures	A	B	C	T
<chem>O=P(O)O</chem>			0	1	center(2,3)	5271
<chem>c1ccc2ccccc2c1</chem>			3	2	4	4742
<chem>c1ccsc1</chem>			3	2	4	4334
<chem>N=CN</chem>			1	0	2	4315
<chem>NC(N)=O</chem>			2	1	3	4167
<chem>O=c1cc[nH]c(=O)[nH]1</chem>			7	1	5	4145
<chem>c1ccc2ncccc2c1</chem>			3	2	4	3519
<chem>c1cscn1</chem>			2	3	1	3466
<chem>c1ccc2[nH]cnc2c1</chem>			5	4	6	3462
<chem>c1c[nH]cn1</chem>			3	2	4	2964
<chem>O=[N+][O-]</chem>			1	0	2	2702

Smiles	2D graph	3D structures	A	B	C	T
<chem>O=CNO</chem>			1	0	2	2477
<chem>NC(=O)O</chem>			1	0	2	2438
<chem>O=S=O</chem>			1	0	2	2375
<chem>c1ccc2[nH]ccc2c1</chem>			3	4	2	2301