

Supplementary Materials for “Disentangled Autoencoding Equivariant Diffusion Model for Controlled Generation of 3D Molecules”

Proofs

Proof of MolDiffdAE objective

For simplicity, here we only consider data with one modality. Let $\mathbf{x}$ be a point in the data space $\mathcal{X}$. Ho et al. [1] prove that the diffusion objective is to maximize the ELBO of $\log p(\mathbf{x}_0)$.

$$\mathbb{E}_{q(\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0) \geq \mathbb{E}_{q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)} [\log p_\theta(\mathbf{x}_{0:T}) - \log q(\mathbf{x}_{1:T}|\mathbf{x}_0)] \quad (1)$$

defined with a decreasing noise schedule $\alpha_t \in (0, 1], t \in 1, \dots, T$: where $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{\alpha_t/\alpha_{t-1}}\mathbf{x}_{t-1}, (1 - \alpha_t/\alpha_{t-1})I)$.

When conditioned on $\mathbf{z}$, the negative ELBO becomes:

$$\begin{aligned} & -\mathbb{E}_{q(\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0) \\ & \leq -\mathbb{E}_{q(\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_0)} \log \frac{p_\theta(\mathbf{x}_0, \mathbf{z})}{q(\mathbf{z}|\mathbf{x}_0)} \\ & \leq -\mathbb{E}_{q(\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \log \frac{p_\theta(\mathbf{x}_{0:T}, \mathbf{z})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)q(\mathbf{z}|\mathbf{x}_0)} \\ & = -\mathbb{E}_{q(\mathbf{x}_0)} \left[L_0 + \sum_{t=2}^T L_t + L_T + \text{KL}(q(\mathbf{z}|\mathbf{x}_0)||p(\mathbf{z})) \right], \end{aligned}$$

where

$$\begin{aligned} L_0 &= -\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_0)} \log p_\theta(\mathbf{x}_0|\mathbf{x}_1, \mathbf{z}) \\ L_T &= \text{KL}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T)) \\ L_t &= \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} \mathbb{E}_{q(\mathbf{z}|\mathbf{x}_0)} \text{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{z})). \end{aligned}$$

We ignore the constant L_T and the boundary case L_0 . Then the RHS can be simplified into:

$$\begin{aligned}
& - \sum_{t=1}^T \mathbb{E}_{\mathbf{x}_0, \mathbf{z}, \mathbf{x}_t} \|\hat{\mathbf{x}}_{0,t}(\mathbf{x}_t, \mathbf{z}, t) - \mathbf{x}_0\|_2^2 \\
& - \mathbb{E}_{q(\mathbf{x}_0)} \text{KL}(q(\mathbf{z}|\mathbf{x}_0)||p(\mathbf{z})) \\
& = -L_D(\theta) - \mathbb{E}_{q(\mathbf{x}_0)} \text{KL}(q(\mathbf{z}|\mathbf{x}_0)||p(\mathbf{z})) \\
& := L_{\text{ELBO}}.
\end{aligned} \tag{2}$$

where $\hat{\mathbf{x}}_{0,t}(\mathbf{x}_t, \mathbf{z}, t)$ is the estimate of the clean data $\mathbf{x}_0$ based on the noisy data $\mathbf{x}_t$ at time t and the embedding $\mathbf{z}$. $L_D(\theta)$ is the same as Eq. 8. This is almost the same form as the DDPM objective in [1], except for the condition $\mathbf{z}$ and the KL divergence term.

By adding a KL divergence regularizer between $q(\mathbf{z})$ and $p(\mathbf{z})$ to the diffusion loss:

$$\begin{aligned}
L &= L_D(\theta) + \beta \text{KL}(q(\mathbf{z})||p(\mathbf{z})) \\
&= L_D(\theta) + \beta [\mathbb{E}_{q(\mathbf{x}_0)} \text{KL}(q(\mathbf{z}|\mathbf{x}_0)||p(\mathbf{z})) - \text{MI}(\mathbf{x}_0, \mathbf{z})],
\end{aligned}$$

when $\beta = 1$:

$$\begin{aligned}
L &= L_D(\theta) + \mathbb{E}_{q(\mathbf{x}_0)} \text{KL}(q(\mathbf{z}|\mathbf{x}_0)||p(\mathbf{z})) - \text{MI}(\mathbf{x}_0, \mathbf{z}) \\
&= -L_{\text{ELBO}} - \text{MI}(\mathbf{x}_0, \mathbf{z}).
\end{aligned} \tag{3}$$

Proof of the closed form solution for the linear manipulation

For simplicity, we omit the bias term $\mathbf{b}$ and the subscript p . The linear manipulation can be formulated as a constrained optimization problem:

$$\min_{\mathbf{z}'} \|\mathbf{z}' - \mathbf{z}\|_2^2$$

subject to $\mathbf{y}' = \mathbf{W}\mathbf{z}'$, where $\mathbf{z}$ satisfies $\mathbf{y} = \mathbf{W}\mathbf{z}$.

Let $\mathbf{W}^+$ be the pseudoinverse of $\mathbf{W}$. Then:

$$\begin{aligned}
\mathbf{z} &= \mathbf{W}^+ \mathbf{W}\mathbf{z} + (\mathbf{I} - \mathbf{W}^+ \mathbf{W})\mathbf{z} \\
&:= \mathbf{z}_r + \mathbf{z}_n,
\end{aligned}$$

where $\mathbf{z}_r \in \mathcal{R}(\mathbf{W})$ and $\mathbf{z}_n \in \text{Null}(\mathbf{W})$ are the row space and null space components, respectively.

Similarly, $\mathbf{z}' := \mathbf{z}'_r + \mathbf{z}'_n$ where $\mathbf{z}'_r \in \mathcal{R}(\mathbf{W})$ and $\mathbf{z}'_n \in \text{Null}(\mathbf{W})$. To satisfy $\mathbf{y}' = \mathbf{W}\mathbf{z}'$, $\mathbf{z}$ must contain a fixed row space component $\mathbf{z}'_r = \mathbf{W}^+ \mathbf{y}'$.

Since $\mathcal{R}(\mathbf{W}) \perp \text{Null}(\mathbf{W})$,

$$\begin{aligned}
\|\mathbf{z}' - \mathbf{z}\|_2^2 &= \|(\mathbf{W}^+ \mathbf{y}' - \mathbf{W}^+ \mathbf{y}) + (\mathbf{z}'_n - \mathbf{z}_n)\|_2^2 \\
&= \|\mathbf{W}^+ \mathbf{y}' - \mathbf{W}^+ \mathbf{y}\|_2^2 + \|\mathbf{z}'_n - \mathbf{z}_n\|_2^2.
\end{aligned}$$

Since the first term is a fixed value, $\|z' - z\|^2$ is minimized when the null space components z'_n and z_n are the same. Thus:

$$z' = W^+y' + (I - W^+W)z = z + W^+(y' - Wz).$$

Extended experiments

Model and training details

The diffusion decoder has 6 EGNN blocks and the encoder has 3. For both networks, the atom type feature is scaled by 0.25 and embedded into a 256-dimensional continuous vector, and edge type feature scaled by 0.125 and embedded into a 64-dimensional continuous vector. The semantic embedding has 32 dimensions. The Wasserstein loss coefficient β in Eq. 1 is 1.0 for the pre-training and 10.0 for the fine-tuning.

For the training of all diffusion models, we use the same noise schedule as in MolDiff [2] with denoising steps $T = 1000$. The MolDiffdAE model is trained with 220,000 steps, and the basic and conditioned diffusion 110,000 steps. The fine-tuning of MolDiffdAE takes another 50,000 steps. The pre-training takes 48 hours on 1 RTX A6000. Sampling one batch of molecules (1 32, with 500~1000 denoising steps) takes around 2 minutes.

Disentanglement analysis

We use the DCI (disentanglement, completeness, informativeness) metric [3]. Specifically, these metrics indicate different preferred aspects of the feature (semantic) space:

- Disentanglement: each dimension should only predict a small number of properties.
- Completeness: each property should only be encoded by a small number of dimensions.
- Informativeness: the properties should be predicted accurately. The original definition based on accuracy is designed for discrete properties. As our properties are continuous, we use the R^2 score as the basis for the informativeness measure.

The DCI scores for the pre-trained and fine-tuned models are shown in Table S2. The fine-tuned model achieves superior performance in all three metrics, especially for the fine-tuned properties (QED, SAS, LogP).

Extended evaluation metrics

Following [2], we use the Jensen-Shannon divergence between the distribution of bond length, bond angles and dihedral angles between the set of the generated molecules and that of the training set. For each category, the JS divergence of each existing bond type (or pair in case of dihedral angles) in the molecule set is calculated separately then averaged across all types. We use randomly sampled coordinates from a Gaussian distribution as the negative control ("random coord").

We also use GFN2-xTB [4] to evaluate the energy of the generated molecules. We use the default parameters for the calculation and invalid molecules are skipped.

Control with randomly-sampled novel semantics

We randomly sample 100 $\mathbf{z}$ ’s from the latent space distribution $\mathcal{N}(0, I)$ and use them to generate 10 novel molecules each. We use the same linear regression models as used in the property manipulation to predict the property profiles of the random $\mathbf{z}$ ’s. These predicted profiles are then used as the objectives for the classifier guidance. Property MAEs are also calculated with respect to the predicted profiles.

Evaluations are the same as in Methods, except that here we use the first molecule generated for each $\mathbf{z}$ as the reference to calculate the Tanimoto similarities. Results are shown in Table S3.

Conformer generation

We could adapt our framework for conformer generation by treating it as an inpainting problem: fixing the atom and bond types and only denoising the atom coordinates. Our empirical results show that using the original embedding of the source conformer and initializing with the source atom coordinates (MolDiffdAE (coord init)), rather than full random Gaussian noise, improves the generation.

We compare it with several baselines based on diffusion models without the semantic guidance:

- DM (coord init): We use the same inpainting problem setup, initialized with the coordinates of the source conformer, using a basic diffusion model without the semantic guidance.
- DM (random init): We use the same basic diffusion model but with random coordinate initialization.
- Random coord: We randomly sample the coordinates from a Gaussian distribution.

The generated conformers are evaluated by the same criteria as in “Extended evaluation metrics”. It is noteworthy that none of the models were specifically trained for conformer generation. With the semantic guidance, MolDiffdAE achieves marginally better results than the basic diffusion model in terms of both the geometric features (Table S5) and conformer energy (Fig. S5B). Example conformers are shown in Fig. S5C. We believe a fine-tuning of the model specifically on conformer data would further improve the performance.

Single-objective manipulation results

We experiment with 10 manipulation tasks on 6 properties. For QED, SAS and LogP, we only test the usually desired directions in drug design (higher QED, lower SAS and LogP). We aim to manipulate molecules to achieve higher (“+” tasks) or lower (“−” tasks) property values and report the success cases where (1) the generated molecules are valid and novel; (2) the property value is improved by at least $0.5\sigma_{p_i}$ where σ_{p_i} is the standard deviation of the property p_i ; (3) the generated molecules have > 0.5 3D shape Tanimoto similarity with the source molecule. We observe that the manipulated molecules and the respective source molecules are generally similar except for the manipulated property (Fig. S12B). Shown in Table S6, MolDiffdAE-based methods even when using simple linear manipulation achieve competitive and

sometimes superior performance compared to classifier guidance. The performance on the more complex properties (QED, SAS and LogP) can be further enhanced by fine-tuning and back-propagation manipulation.

Ablation study

We compare our pre-trained model **MolDiffdAE** with two different settings:

- **MolDiffdAE-64**: the dimension of the latent space is 64.
- **MolDiffdAE-Wass-**: the weight for the Wasserstein loss is 0.

For single-property manipulation tasks, their performance is generally worse compared to the pre-trained model. Specifically, **MolDiffdAE-Wass-** had near-zero success rates, while **MolDiffdAE-64** is highly biased towards some properties, achieving decent performance, but has much worse performance on other properties. We further observe that without the Wasserstein loss, the disentanglement in latent space is poor. On the other hand, for higher-dimensional embeddings, enforcing disentanglement through the Wasserstein loss is more challenging. As shown in Fig. S6, the latent spaces of the latter two settings are less disentangled, where strong correlations exist between dimensions. These results highlight the importance of the regularization. From another perspective, it is possible to design and incorporate regularizers that are more effective on higher dimensions, achieving both higher expressivity and better disentanglement of the latent space.

Table. S1 The list of properties used for the property preservation analysis.

2D		
NHOHCount	fr_N_O	fr_ketone.Topliss
NOCCount	fr_Ndealkylation1	fr_lactam
NumAliphaticCarbocycles	fr_Ndealkylation2	fr_lactone
NumAliphaticHeterocycles	fr_Nhpyrrole	fr_methoxy
NumAliphaticRings	fr_SH	fr_morpholine
NumAromaticCarbocycles	fr_aldehyde	fr_nitrile
NumAromaticHeterocycles	fr_alkyl_carbamate	fr_nitro
NumAromaticRings	fr_alkyl_halide	fr_nitro_arom
NumHAcceptors	fr_allylic_oxid	fr_nitro_arom_nonortho
NumHDonors	fr_amide	fr_nitroso
NumHeteroatoms	fr_amidine	fr_oxazole
NumRotatableBonds	fr_aniline	fr_oxime
NumSaturatedCarbocycles	fr_aryl_methyl	fr_para_hydroxylation
NumSaturatedHeterocycles	fr_azide	fr_phenol
NumSaturatedRings	fr_azo	fr_phenol_noOrthoHbond
RingCount	fr_barbitur	fr_phos_acid
fr_Al_COO	fr_benzene	fr_phos_ester
fr_Al_OH	fr_benzodiazepine	fr_piperdine
fr_Al_OH_noTert	fr_bicyclic	fr_piperzine
fr_ArN	fr_diazo	fr_priamide
fr_Ar_COO	fr_dihydropyridine	fr_prisulfonamd
fr_Ar_N	fr_epoxide	fr_pyridine
fr_Ar_NH	fr_ester	fr_quatN
fr_Ar_OH	fr_ether	fr_sulfide
fr_COO	fr_furan	fr_sulfonamd
fr_COO2	fr_guanido	fr_sulfone
fr_C_O	fr_halogen	fr_term_acetylene
fr_C_O_noCOO	fr_hdrzine	fr_tetrazole
fr_C_S	fr_hdrzone	fr_thiazole
fr_HOCCN	fr_imidazole	fr_thiocyan
fr_Imin	fr_imide	fr_thiophene
fr_NH0	fr_isocyan	fr_unbrch_alkane
fr_NH1	fr_isothiocyan	fr_urea
fr_NH2	fr_ketone	
3D		
PMI1	NPR2	Asphericity
PMI2	RadiusOfGyration	SphericityIndex
PMI3	InertialShapeFactor	PBF
NPR1	Eccentricity	

Table. S2 Disentanglement analysis of the latent semantic space

	MolDiffdAE	MolDiffdAE-ft
Fine-tuned properties		
Disentanglement $\uparrow$	0.21	0.36
Completeness $\uparrow$	0.12	0.24
Informativeness $\uparrow$	-5.42	0.24
Fig. 2 properties		
Disentanglement $\uparrow$	0.25	0.28
Completeness $\uparrow$	0.20	0.22
Informativeness $\uparrow$	-1.33	0.58

Table. S3 Property mean absolute error (MAE) and molecular similarity of generated molecules from random novel semantics. We report the MAEs aggregated from the selected 2D ("All 2D") and 3D ("All 3D") property groups and their combination ("All"), where all properties are standardized into z-scores. For the individual properties, the MAEs of raw values are reported. Property groups are in italics. The best values are in bold. (VAE: variational autoencoder)

	MolDiffdAE	Classifier Guidance	Random
Property MAE $\downarrow$			
<i>All</i> (standardized)	0.59$\pm$0.15	0.69 $\pm$ 0.15	1.13 $\pm$ 0.09
<i>All 2D</i> (standardized)	0.62$\pm$0.17	0.65 $\pm$ 0.15	1.13 $\pm$ 0.10
MolWt	11.00$\pm$7.76	16.63 $\pm$ 12.9	87.77 $\pm$ 66.07
NOCount	0.87$\pm$0.73	1.72 $\pm$ 1.28	2.31 $\pm$ 1.73
fr_halogen	0.47$\pm$0.38	0.77 $\pm$ 0.67	0.97 $\pm$ 0.74
fr_C_O	0.65$\pm$0.48	0.78 $\pm$ 0.62	1.02 $\pm$ 0.77
NumAromaticRings	0.91 $\pm$ 0.66	0.90$\pm$0.65	1.10 $\pm$ 0.83
RingCount	0.73$\pm$0.59	0.82 $\pm$ 0.64	1.13 $\pm$ 0.85
QED	0.13$\pm$0.10	0.13 $\pm$ 0.10	0.19 $\pm$ 0.14
SAS	1.11 $\pm$ 0.86	1.01$\pm$0.81	0.88 $\pm$ 0.66
LogP	0.93$\pm$0.75	1.30 $\pm$ 1.00	1.56 $\pm$ 1.17
...	...	...	...
<i>All 3D</i> (standardized)	0.38$\pm$0.16	0.92 $\pm$ 0.51	1.13 $\pm$ 0.25
Asphericity	0.06$\pm$0.04	0.22 $\pm$ 0.17	0.22 $\pm$ 0.17
PBF	0.13$\pm$0.10	0.27 $\pm$ 0.20	0.35 $\pm$ 0.26
RadiusOfGyration	0.16$\pm$0.12	0.57 $\pm$ 0.44	0.67 $\pm$ 0.50
...	...	...	...
Similarity $\uparrow$			
Tanimoto 2D	0.30$\pm$0.11	0.27 $\pm$ 0.09	0.26 $\pm$ 0.08
Tanimoto 3D	0.50$\pm$0.10	0.39 $\pm$ 0.06	0.38 $\pm$ 0.07

Table. S4 Multi-property manipulation results, evaluated by Jensen-Shannon divergence between the bond length, bond angle, and dihedral angle distribution between the generated conformers and the source dataset.

	Bond length ↓	Bond angle ↓	Dihedral angle ↓
Random coord	0.52	0.56	0.41
<i>Random generation</i>			
DM random	0.43	0.30	0.14
MolDiffdAE	0.41	0.35	0.32
<i>Manipulation</i>			
Classifier Guidance	0.41	0.30	0.29
MolDiffdAE+lm	0.44	0.37	0.33
MolDiffdAE-ft+lm	0.41	0.29	0.27
MolDiffdAE-ft+bp	0.38	0.26	0.26
MolDiffdAE-ft+RAG-interp	0.38	<u>0.23</u>	<u>0.17</u>
MolDiffdAE-ft+RAG-reopt	0.38	0.24	<u>0.17</u>

Table. S5 Conformer generation results, evaluated by Jensen-Shannon divergence between the bond length, bond angle, and dihedral angle distribution between the generated conformers and the source dataset. (init: initialization; DM: diffusion model)

	Bond length ↓	Bond angle ↓	Dihedral angle ↓
Random coord	0.52	0.56	0.41
DM (random init)	0.52	0.55	0.39
DM (coord init)	0.50	0.45	0.28
MolDiffdAE (coord init)	0.47	0.40	0.25

Table. S6 Single objective manipulation results by success rate (improvement threshold = $0.5\sigma_{p_i}$). (VAE: variational autoencoder; QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

GEOM			
	QED+	SAS−	LogP−
Random Diffusion	0.02	0.04	0.04
Random MolDiffdAE	0.17	0.08	0.15
Conditional Diffusion	0.07	0.04	0.34
VAE (2D)+lm	0.03	0.06	0.03
Classifier Guidance	0.27	0.41	0.79
MolDiffdAE+lm	0.30	0.07	0.62
MolDiffdAE-ft+lm	0.43	0.30	0.73
MolDiffdAE-ft+bp	<u>0.35</u>	<u>0.32</u>	<u>0.75</u>
	Halogen+	Halogen−	
Random Diffusion	0.00	0.05	
Random MolDiffdAE	0.04	0.33	
Conditional Diffusion	0.19	0.14	
VAE (2D)+lm	0.03	0.06	
Classifier Guidance	<u>0.73</u>	0.96	
MolDiffdAE+lm	0.53	<u>0.92</u>	
MolDiffdAE-ft+lm	0.75	0.89	
CrossDocked			
	QED+	SAS−	LogP−
Random Diffusion	0.06	0.05	0.04
Random MolDiffdAE	0.14	0.14	0.08
Conditional Diffusion	0.04	0.01	0.19
VAE (2D)+lm	0.03	0.04	0.05
Classifier Guidance	0.22	<u>0.25</u>	0.77
MolDiffdAE+lm	0.04	0.05	0.35
MolDiffdAE-ft+lm	<u>0.15</u>	0.26	<u>0.65</u>
MolDiffdAE-ft+bp	0.10	0.17	0.50

Table. S7 Multi objective manipulation results by success rate with only the property improvement constraints (improvements $> 0.5\sigma_{p_i}$) and no similarity constraints. (QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient; Asp: asphericity; RoG: radius of gyration)

	MolDiffdAE+lm	MolDiffdAE-ft+lm	Classifier Guidance	VAE (2D)+lm
QED+,SAS-	0.05	0.21	0.02	0.23
QED+,LogP-	0.08	0.29	0.17	0.09
LogP-,SAS-	0.00	0.13	0.02	0.10
QED+,Halogen-	0.25	0.48	0.15	0.24
LogP-,Halogen-	0.68	0.73	0.73	0.07
SAS-,Halogen-	0.09	0.50	0.05	0.25
QED+,Halogen+	0.24	0.33	0.18	0.05
LogP-,Halogen+	0.45	0.44	0.04	0.01
SAS-,Halogen+	0.14	0.33	0.08	0.11
QED+,SAS-,Halogen+	0.03	0.15	0.08	0.08
QED+,LogP-,Halogen+	0.01	0.14	0.00	0.03
LogP-,SAS-,Halogen+	0.00	0.06	0.00	0.01
QED+,SAS-,LogP-	0.00	0.13	0.01	0.14
QED+,SAS-,Halogen-	0.05	0.10	0.03	0.17
QED+,LogP-,Halogen-	0.00	0.25	0.05	0.11
LogP-,SAS-,Halogen-	0.00	0.23	0.08	0.12
Asp+,RoG+	0.64	0.78	0.02	0.59
Asp+,RoG-	0.00	0.25	0.00	0.00
Asp-,RoG+	0.51	0.21	0.00	0.04
Asp-,RoG-	0.71	0.64	0.00	0.09

Table. S8 Ablation study results by success rate (improvement threshold = $0.5\sigma_{p_i}$). (Asp: asphericity; RoG: radius of gyration)

	Asp+	RoG+	Halogen+
MolDiffdAE (base)	<u>0.21</u>	0.44	0.53
MolDiffdAE-64	0.36	<u>0.04</u>	0.00
MolDiffdAE-Wass-	0.00	0.00	0.00
	Asp-	RoG-	Halogen-
MolDiffdAE (base)	<u>0.47</u>	0.44	0.92
MolDiffdAE-64	0.58	<u>0.41</u>	0.04
MolDiffdAE-Wass-	0.00	0.00	0.00
	QED+	SAS-	LogP-
MolDiffdAE (base)	0.30	<u>0.07</u>	0.62
MolDiffdAE-64	<u>0.16</u>	0.13	<u>0.09</u>
MolDiffdAE-Wass-	0.05	<u>0.07</u>	0.02

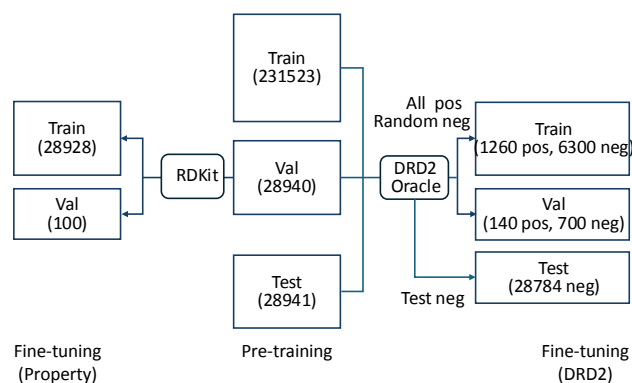


Fig. S1 Construction of the fine-tuning data

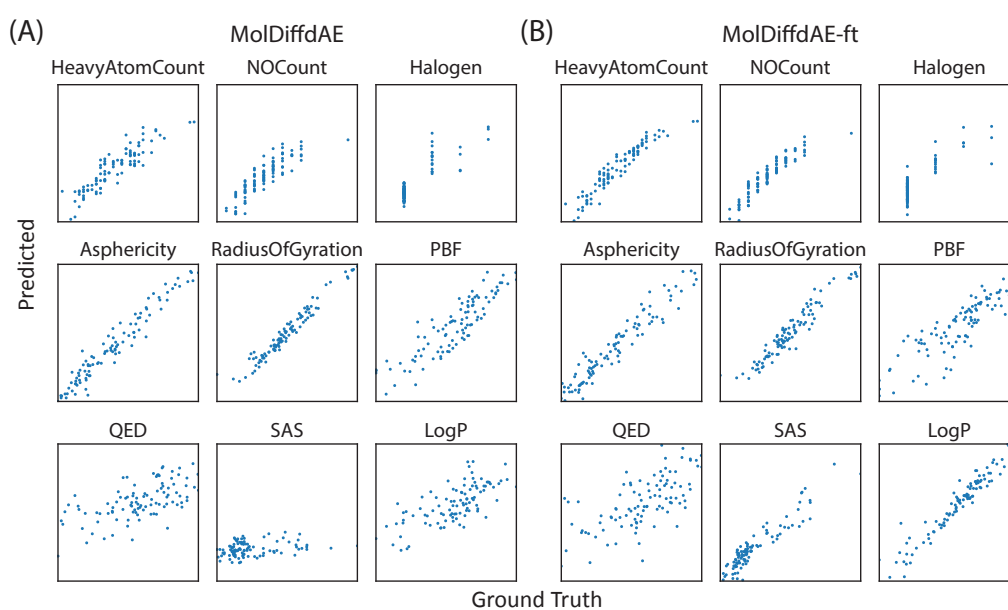


Fig. S2 Linear regression performance of the (A) pre-trained and (B) fine-tuned model. The x-axis is the true value and the y-axis is the predicted value.

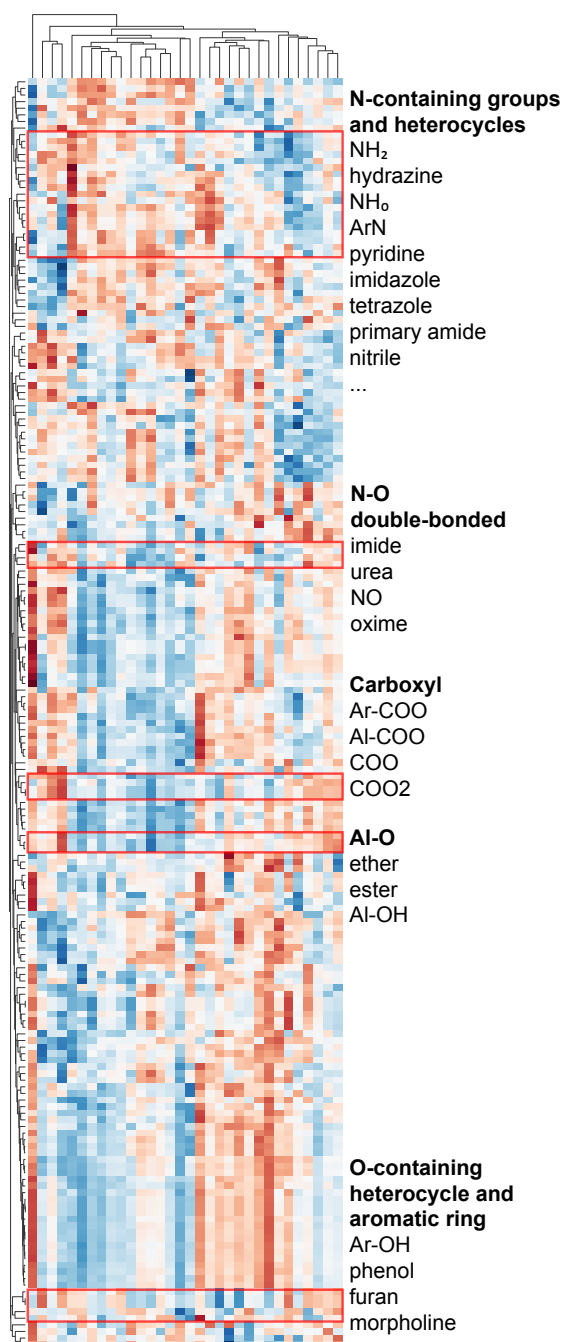


Fig. S3 Linear regression coefficient of the embedding of the pre-trained model with respect to all the RDKit properties. For each property (row), the weights are scaled to unit variance. We annotate some of the clusters by their property contents.

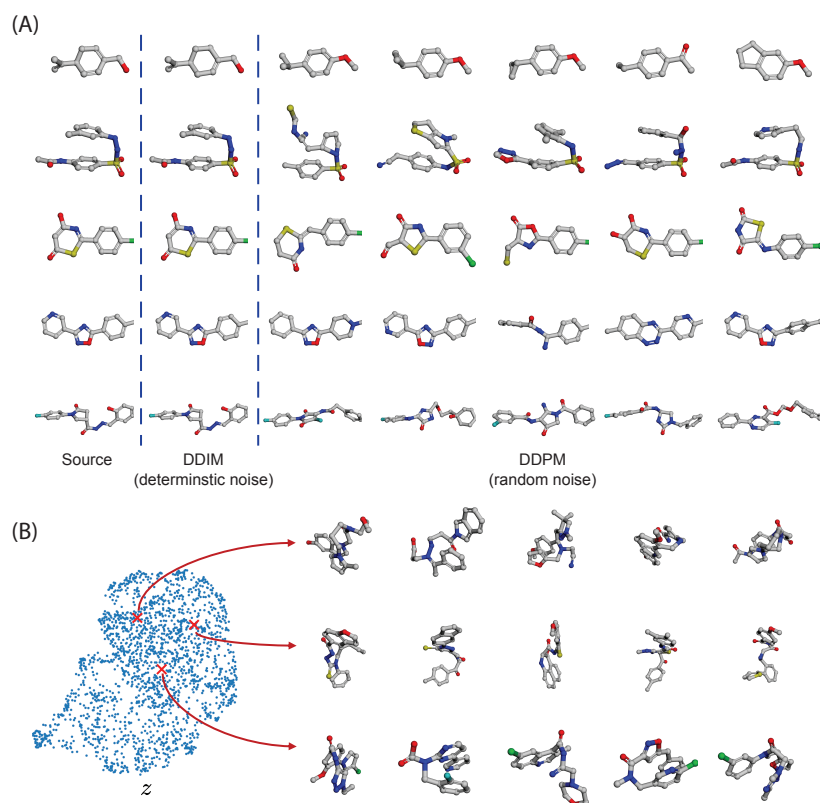


Fig. S4 Semantics-guided 3D molecule generation results. (A) Example of semantics-guided generation results from the semantic embedding of the source molecules. DDIM (denoising diffusion implicit model; deterministic) leads to near-perfect generation, while DDPM (denoising diffusion probabilistic model; stochastic) produces molecules strongly controlled by the embedding; (B) Example of semantics-guided generated molecules from randomly sampled semantic embeddings. Molecules in the same row are generated from the same embedding.

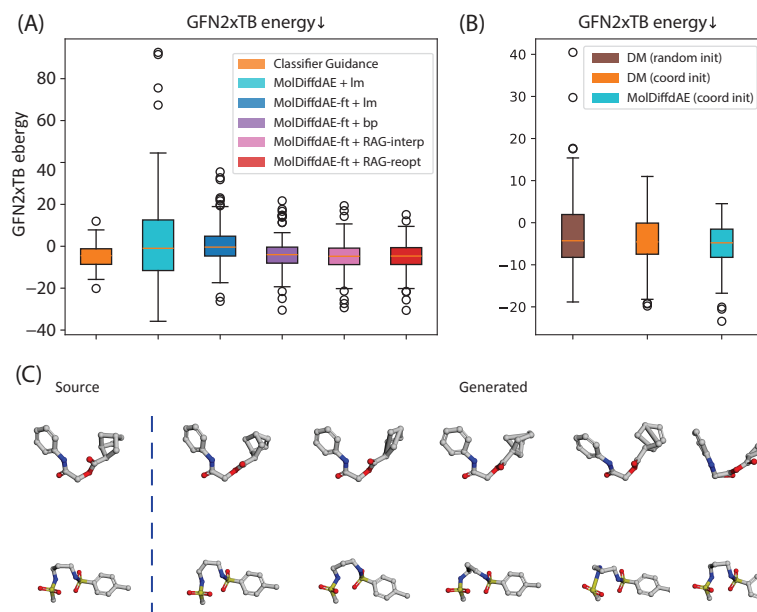


Fig. S5 GFN2-xTB energy analysis of (A) multiple property manipulation results and (B) conformer generation results (note that the energy for the randomly sampled coordinates are too large, so is not plotted here; whiskers indicate minimum and maximum). (C) Example generated conformers using MolDiffdAE. (DM: diffusion model)

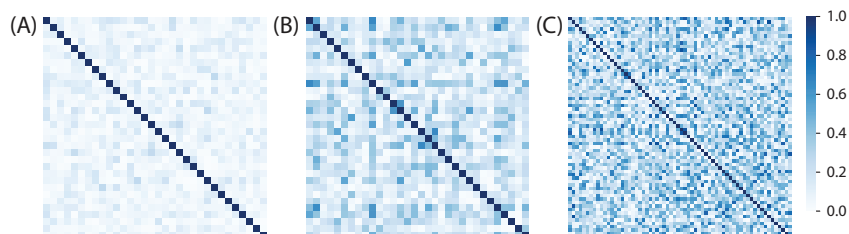


Fig. S6 Correlation between the dimensions of the semantic space for (A) the pre-trained 32-dimensional model, (B) the 32-dimensional model without Wasserstein loss, and (C) the 64-dimensional model.

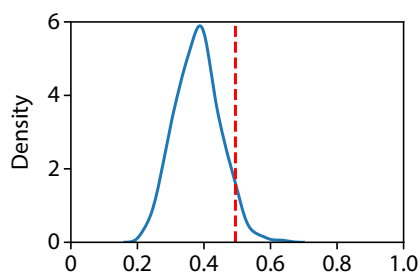


Fig. S7 Distribution of the 3D Tanimoto similarity score from 1000 random pairs in the validation set. The dashed line is the 95% quantile.

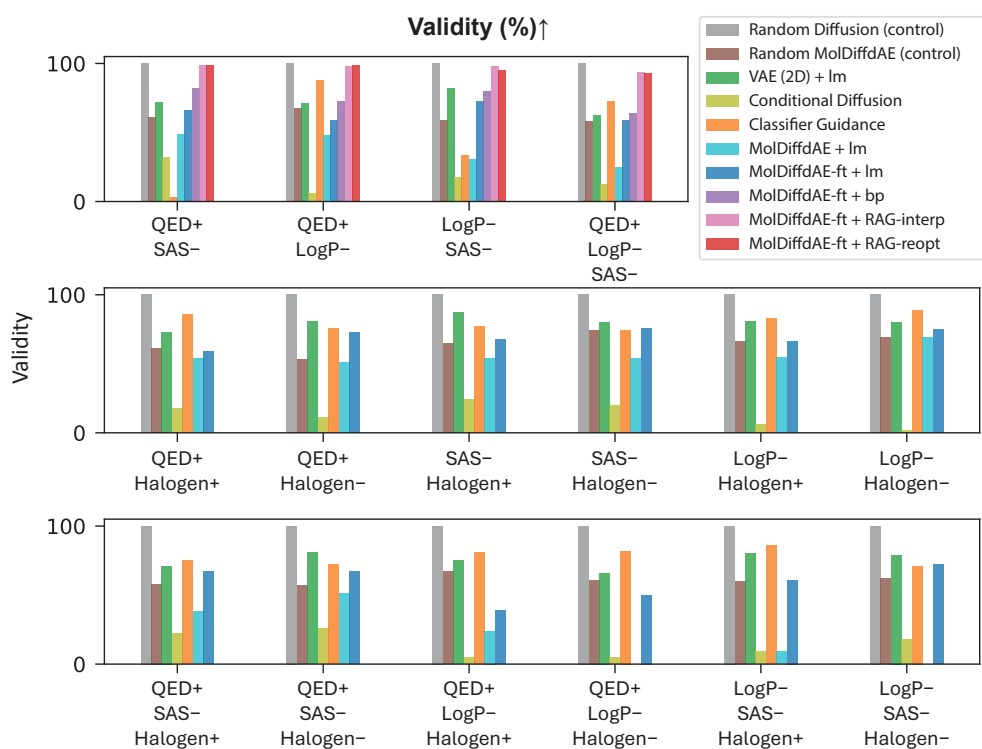


Fig. S8 Multi-objective manipulation results by molecular validity. (VAE: variational autoencoder; QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

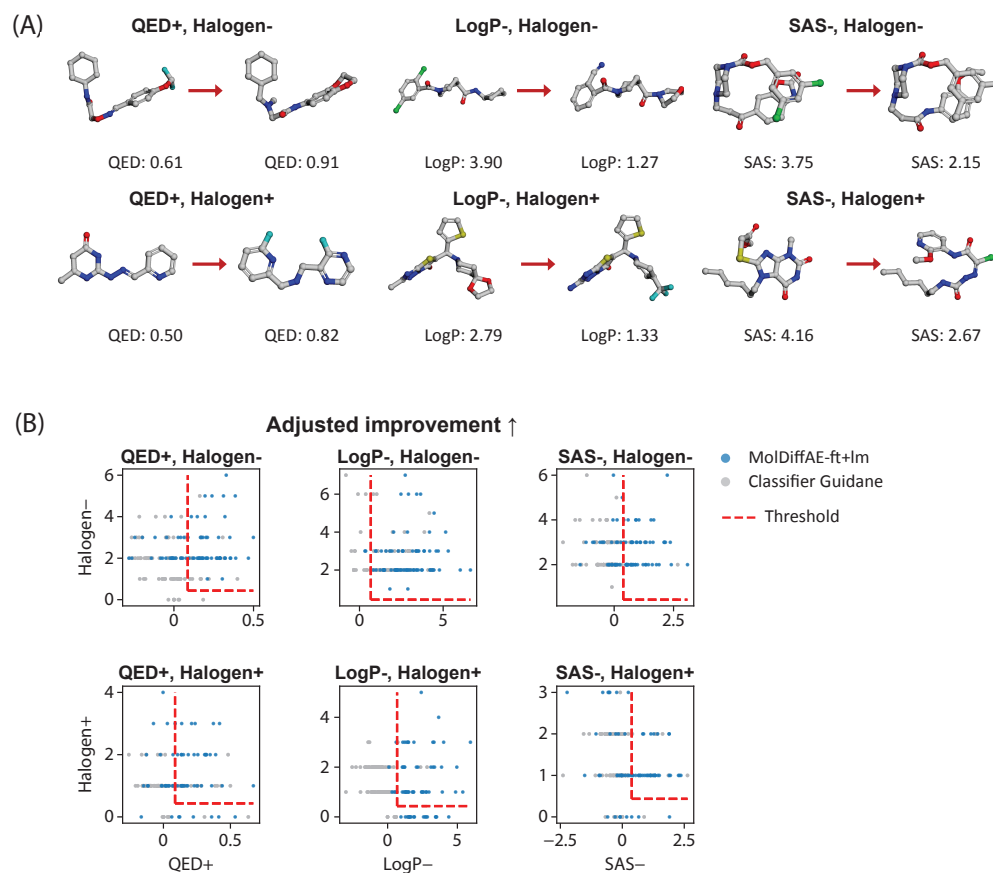


Fig. S9 Results of 2-objective manipulation. (A) Examples of manipulated molecules (right) compared to sources (left), with the values of the target properties. (B) Improvement of both target properties for the manipulated molecules. Only molecules that are valid and have > 0.5 3D Tanimoto similarity with the source are plotted. Improvements are adjusted by the signs of the objectives (“+” or “-”) so that more positive values are always preferred. For tasks involving only QED, SAS and LogP, we show results using back-propagation manipulation. For the other tasks, we use linear manipulation. (QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

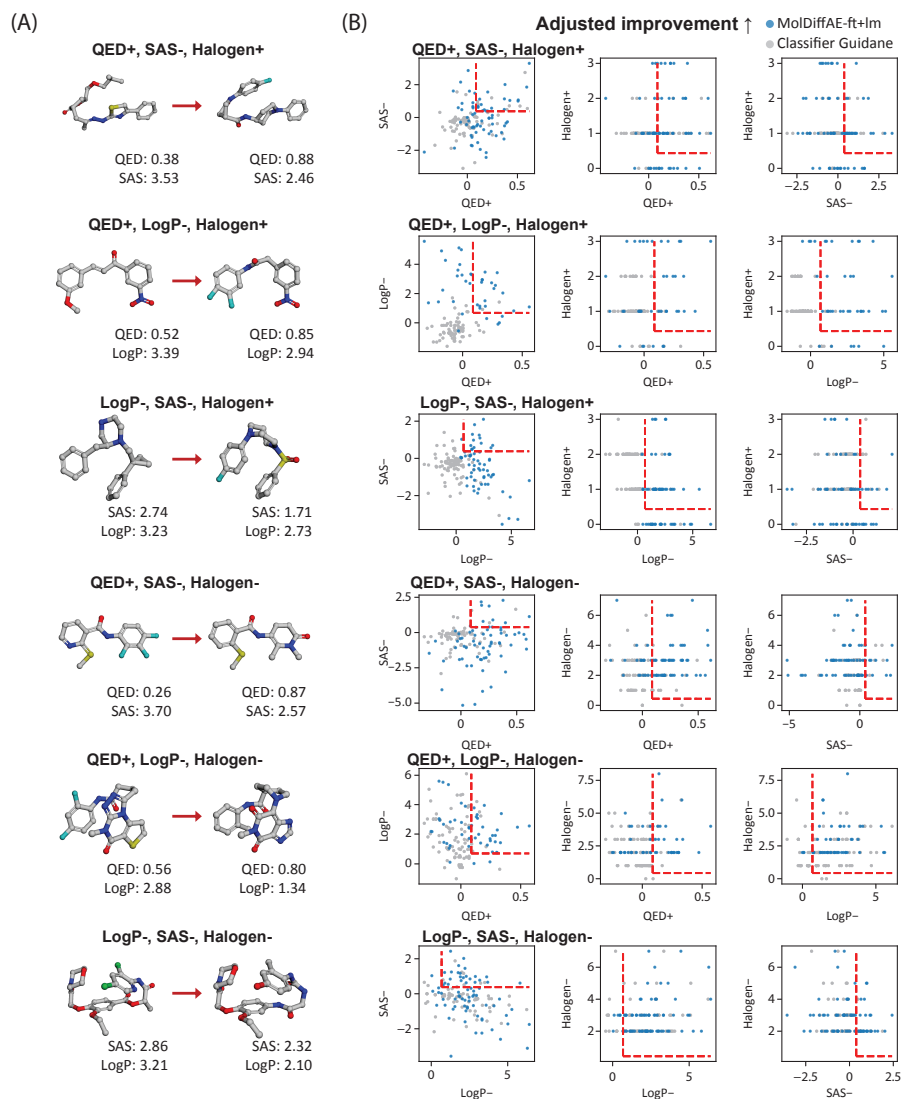


Fig. S10 Results of 3-objective manipulation. (A) Examples of manipulated molecules (right) compared to sources (left), with the values of the target properties. (B) Pairwise comparison of the improvement of the target properties for the manipulated molecules. Only molecules that are valid and have > 0.5 3D Tanimoto similarity with the source are plotted. For (QED+, SAS-, LogP-), we show results using back-propagation manipulation. For the other tasks, we use linear manipulation. (QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

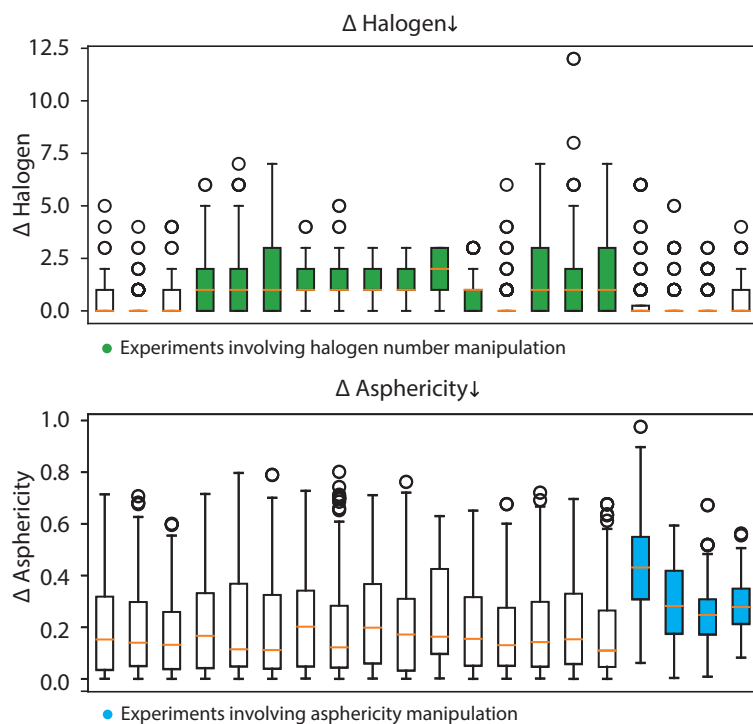


Fig. S11 Change of selected property values upon property manipulation experiments. For each experiment, we calculate the difference of halogen count (top) and asphericity (bottom) between the manipulated molecule and that of the source molecule. Each box represents the results of one experiment with a specified property target. Experiments that explicitly manipulated the respective properties (top: halogen; bottom: asphericity) are highlighted in green. In general, the value of the property in question only drastically changes when it is directly manipulated, while staying mostly the same when manipulating other properties (whiskers indicate minimum and maximum).

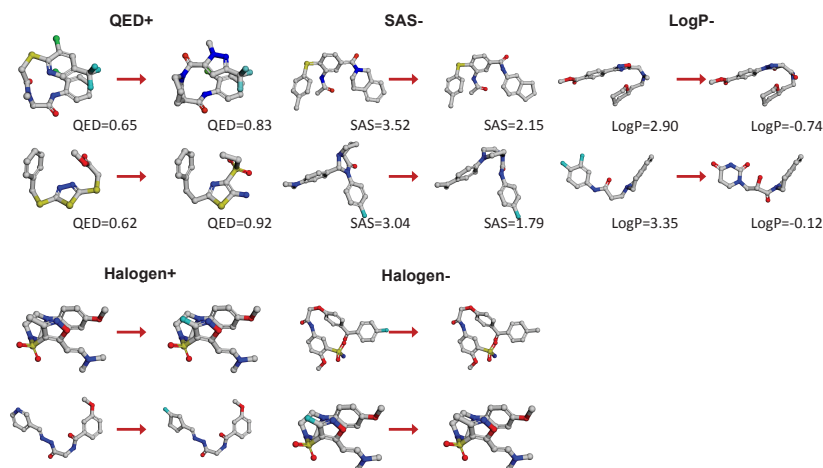


Fig. S12 Sample single-objective manipulation results. For tasks involving QED, SAS and LogP, we show results using back-propagation manipulation. For the other tasks, we use linear manipulation. (QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

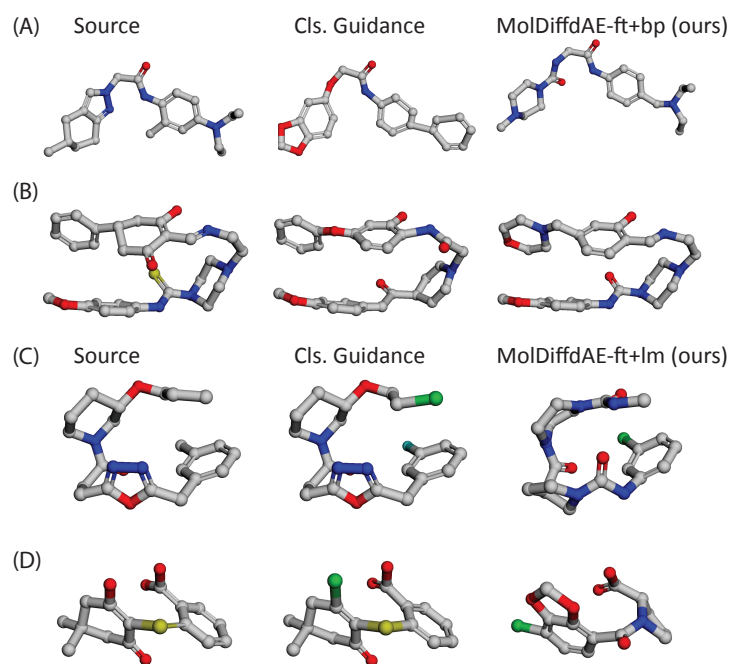


Fig. S13 3D conformation of the source and manipulated molecules in Fig. 4.

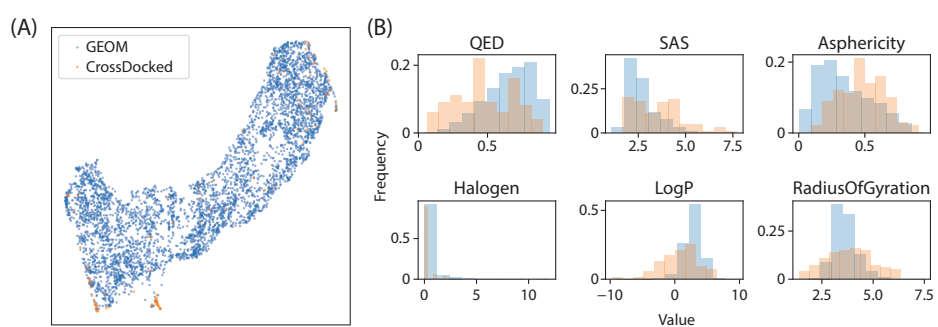


Fig. S14 (A) CrossDocked samples in the latent space of the pre-trained model; (B) Distribution of the selected properties for GEOM and CrossDocked test sets. (QED: quantitative estimate of drug-likeness; SAS: synthetic accessibility score; LogP: n-octanol/water partition coefficient)

References

- [1] Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020).
- [2] Peng, X., Guan, J., Liu, Q. & Ma, J. Moldiff: Addressing the atom-bond inconsistency problem in 3d molecule diffusion generation. *arXiv preprint arXiv:2305.07508* (2023).
- [3] Eastwood, C. & Williams, C. K. *A framework for the quantitative evaluation of disentangled representations* (2018).
- [4] Bannwarth, C., Ehlert, S. & Grimme, S. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation* **15**, 1652–1671 (2019).