nature computational science

Supplementary information

https://doi.org/10.1038/s43588-025-00850-5

Pharmacophore-oriented 3D molecular generation toward efficient feature-customized drug discovery

In the format provided by the authors and unedited

Supplementary Information

Content

Section 1 Supplementary Notes

1.1 The overall training and sampling procedure

The algorithms for training and sampling are listed as follows.

Algorithm 1 Training Procedure

Input: Pharmacophore-ligand geometry $\{\mathcal{P}, \mathcal{M}\}_{i=1}^{N}$, graph denoiser $\phi_{\theta 1}$, molecular size module $\varphi_{\theta 2}$

while $\phi_{\theta 1}$ and $\varphi_{\theta 2}$ have not converged do

  Sample diffusion time $t \sim \mathrm{Uniform}(\{0, \ldots, T\})$

  Move the geometry to make the center of mass of pharmacophore zero

  Perturb x_0 to obtain x_t:

  $\mathrm{x}_t = \sqrt{\bar{\alpha}_t}\mathrm{x}_0 + (1 - \bar{\alpha}_t)\epsilon$, where $\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$

  Perturb v_0 to obtain v_t:

  $\log c = \log(\bar{\alpha}_t \mathrm{v}_0 + (1 - \bar{\alpha}_t)/K)$

  $\mathrm{v}_t = \mathrm{one\_hot}(\arg\max_i [g_i + \log c_i])$, where $g \sim \mathrm{Gumbel}(0, 1)$

  Perturb b_0 to obtain b_t:

  $\log c' = \log(\bar{\alpha}_t \mathrm{b}_0 + (1 - \bar{\alpha}_t)/K')$

  $\mathrm{b}_t = \mathrm{one\_hot}(\arg\max_i [g'_i + \log c'_i])$, where $g' \sim \mathrm{Gumbel}(0, 1)$

  Embed direction matching feature D_{ij} into x_t and v_t

  Predict $[\hat{\mathrm{x}}_0, \hat{\mathrm{v}}_0, \hat{\mathrm{b}}_0]$ with $\phi_{\theta 1}$: $[\hat{\mathrm{x}}_0, \hat{\mathrm{v}}_0, \hat{\mathrm{b}}_0] = \phi_{\theta 1}([\mathrm{x}_t, \mathrm{v}_t, \mathrm{b}_t], t, \mathcal{P})$

  Predict atom count $\widehat{N}_{atoms}$ with $\varphi_{\theta 2}$: $\widehat{N}_{atoms} = \varphi_{\theta 2}(\mathcal{P})$

  Compute the weighted loss $\mathcal{L}$ on atom coordinates, atom types, bond types, and atom count:

  $\mathcal{L} = \|\mathrm{x}_0 - \hat{\mathrm{x}}_0\|^2 + \lambda_1 \mathrm{KL}(c(\mathrm{v}_t, \mathrm{v}_0) \,||\, c(\mathrm{v}_t, \hat{\mathrm{v}}_0)) + \lambda_2 \mathrm{KL}(c(\mathrm{b}_t, \mathrm{b}_0) \,||\, c(\mathrm{b}_t, \hat{\mathrm{b}}_0)) + QD(N_{atoms}, \widehat{N}_{atoms})$

  Jointly update θ_1 and θ_2 by minimizing $\mathcal{L}$

end while

Algorithm 2 Sampling Procedure

Input: The pharmacophore model $\mathcal{P}$, the learned denoising model $\phi_{\theta 1}$, the learned molecular size module $\varphi_{\theta 2}$

Output: Generated ligand molecule $\mathcal{M}$ that aligns to the pharmacophore model

Compute the atom count in $\mathcal{M}$ with $\varphi_{\theta 2}$: $\widehat{N}_{atoms} = \varphi_{\theta 2}(\mathcal{P})$

Move the center of mass of pharmacophore model to zero

Initialize atom coordinates x_T, atom types v_T, and bond types b_T:

$\mathrm{x}_T \sim \mathcal{N}(0, \boldsymbol{I})$

$\mathrm{v}_T = \mathrm{one\_hot}(\arg\max_i g_i)$, where $g \sim \mathrm{Gumbel}(0,1)$

$\mathrm{b}_T = \mathrm{one\_hot}(\arg\max_i g'_i)$, where $g' \sim \mathrm{Gumbel}(0,1)$

for t in $T, T-1, \ldots, 1$ **do**

  Embed direction matching feature D_{ij} into x_t and v_t

  Compute $[\hat{\mathrm{x}}_0, \hat{\mathrm{v}}_0, \hat{\mathrm{b}}_0]$ with $\phi_{\theta 1}$: $[\hat{\mathrm{x}}_0, \hat{\mathrm{v}}_0, \hat{\mathrm{b}}_0] = \phi_{\theta 1}([\mathrm{x}_t, \mathrm{v}_t, \mathrm{b}_t], t, \mathcal{P})$

  Sample x_{t-1} from the posterior $p_\theta(\mathrm{x}_{t-1} \mid \mathrm{x}_t, \hat{\mathrm{x}}_0)$

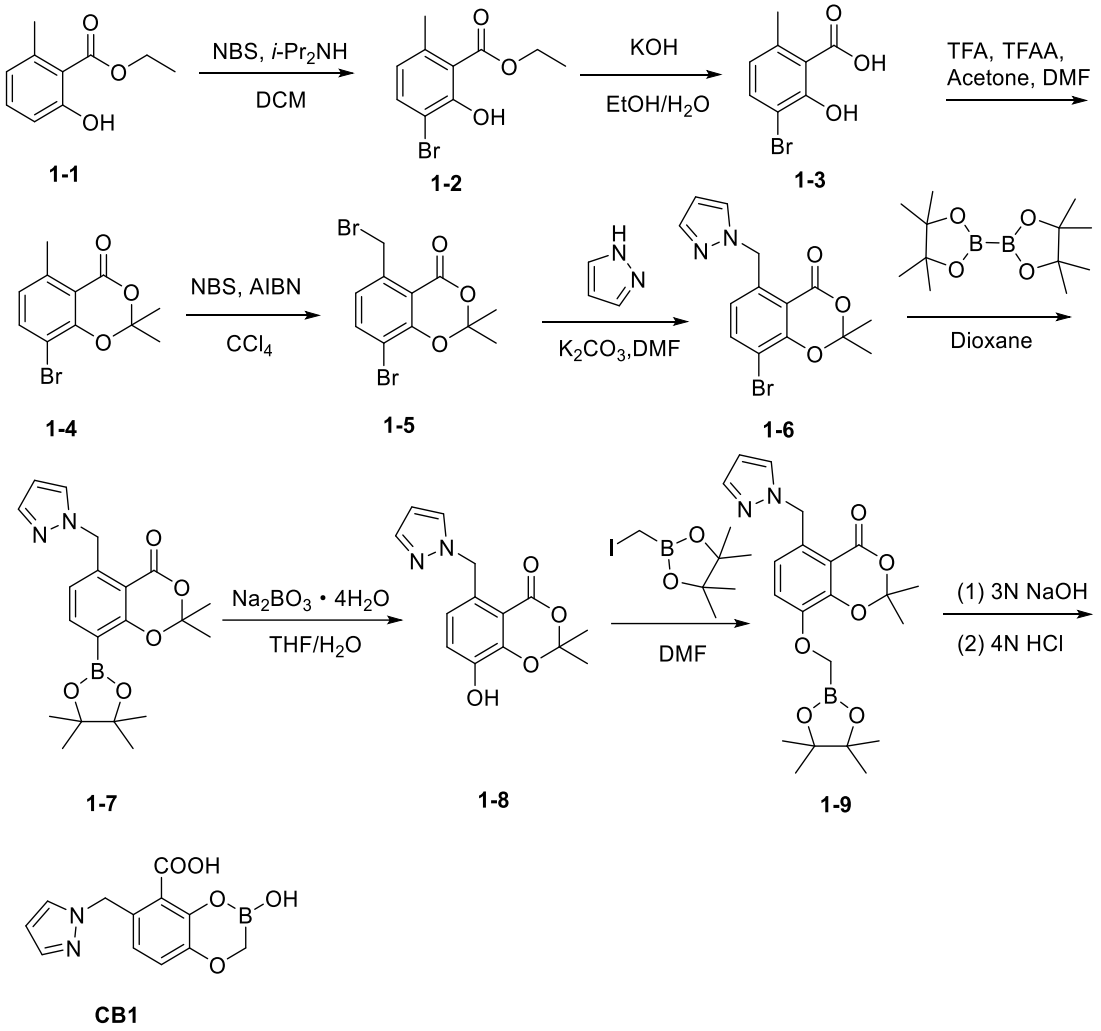
  Sample v_{t-1} from the posterior $p_\theta(\mathrm{v}_{t-1} \mid \mathrm{v}_t, \hat{\mathrm{v}}_0)$

  Sample b_{t-1} from the posterior $p_\theta(\mathrm{b}_{t-1} \mid \mathrm{b}_t, \hat{\mathrm{b}}_0)$

end for

1.2 Chemical synthesis and characterization of CB1 and CB2

Scheme 1. The synthetic route for **CB1**



Synthesis of ethyl 3-bromo-2-hydroxy-6-methylbenzoate (1-2)

The solution of ethyl 2-hydroxy-6-methylbenzoate and DIPA (0.94ml, 0.1eq) in DCM (120ml) was cooled to 0°C. Subsequently, the solution of N-bromosuccinimide (11.8g, 1eq) in DCM (170ml) was added dropwise to the above mixed solution. On completion of the addition, the cooling bath was removed and the mixture was stirred at room temperature overnight. The reaction solution was evaporated to dryness. The residue was purified by silica chromatography to obtain **1-2**(16g, 93% yield).

Synthesis of 3-bromo-2-hydroxy-6-methylbenzoic acid (1-3)

Ethyl 3-bromo-2-hydroxy-6-methylbenzoate (**1-2**) (16g, 1eq) was dissolved in H_2O/EtOH (40ml/60ml) and KOH (16g, 4eq) was added. The resulting solution was heated at 80 °C for 6 h, cooled to room temperature, and then pH was adjusted to 1 with HClconc. Water was added to the reaction mixture and extracted three times with EtOAc. The combined organic layer was washed with brine, dried over anhydrous Na_2SO_4, filtered, and concentrated in vacuo to afford **1-3** (13.5 g, 93.5 % yield).

Synthesis of 8-bromo-2,2,5-trimethyl-4H-benzo[d][1,3]dioxin-4-one (1-4)

In an Ar atmosphere, a mixed solution of trifluoroacetic anhydride (63 mL, 1ml/mmol) and acetone (47 mL, 0.8 ml/mmol) was slowly added to a solution of **1-3** (14.4 g, 62.3 mmol) in TFA (87 mL, 1,4 ml/mmol) and DMF (63 mL, 1ml/mmol) at 0°C. After the dropwise addition, the reaction was heated to 100°C and stirred for 24h. Dilute the residue with EtOAc and saturated aqueous $NaHCO_3$. Extract the mixture with EtOAc. Dry the extracts over Na_2SO_4 and concentrate. Purify the residue by flash column chromatography (hexane/EtOAc = 5/1) to obtain the product **1-4** (12.8g, 75.8% yield).

Synthesis of 8-bromo-5-(bromomethyl)-2,2-dimethyl-4H-benzo[d][1,3] dioxin-4-one (1-5)

A solution of **1-4** (12.8 g, 1 eq) in 410 mL of CCl_4 was refluxed for 4 h with NBS (10.1 g, 1.2 eq) and AIBN (775 mg, 0.1 eq). The solvent was removed under reduced pressure, and the residue purified by silica gel chromatography (hexane: EtOAc = 10:1) to afford **1-5** (9 g, 54% yield).

Synthesis of 5-((1H-pyrazol-1-yl)methyl)-8-bromo-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (1-6)

To **1-5**, pyrazole (600 mg, 1.5 eq) and K_2CO_3 (1.6 g, 2 eq), 50ml anhydrous DMF was added in Ar. The mixture was stirred at room temperature overnight. After completion of the reaction, the mixture was quenched with water, extracted with EA, washed with water and brine, dried over anhydrous sodium sulfate, and the solvent removed under reduced pressure. The crude product was purified by column chromatography to afford **1-6** (1g, 50.1% yield).

Synthesis of 5-((1H-pyrazol-1-yl)methyl)-2,2-dimethyl-8-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-4H-benzo[d][1,3]dioxin-4-one (1-7)

In Ar atmosphere, to a 50 mL two-necked round-bottom flask, add **1-6** (1g, 1eq), bis(pinacolato)-diboron (1.5g, 2eq), Pd(dppf)Cl_2(439mg，0.2eq), AcOK (883mg, 3eq), 30 mL of anhydrous dioxane. The mixture was heated to 80℃ for 24h. Then the water was added to the mixture of reaction. Extract the aqueous layer with EA, wash with brine, dry over anhydrous Na_2SO_4. And the organic layer was removed under reduced pressure to afford crude product.

Synthesis of 5-((1H-pyrazol-1-yl)methyl)-8-hydroxy-2,2-dimethyl-4H-benzo[d][1,3]dioxin-4-one (1-8)

The crude product of **1-7** was dissolved in THF (15ml): H_2O (15ml). Then added the $Na_2BO_3 \cdot 4H_2O$ to the above solution and stirred for 30 min at room temperature. The mixture was quenched with water, extracted with EA, washed with brine, dried over Na_2SO_4, and the solvent removed under reduced pressure. The residue was purified by column chromatography to afford **1-8** (410 mg， two-step reaction yield 50%).

Synthesis of 5-((1H-pyrazol-1-yl)methyl)-2,2-dimethyl-8-((4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)methoxy)-4H-benzo[d][1,3]dioxin-4-one (1-9)

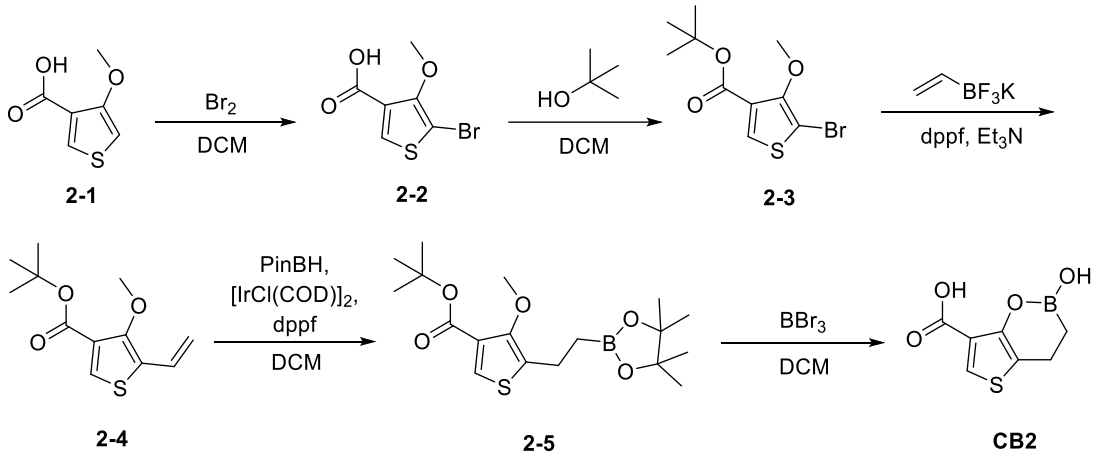
In Ar, the **1-8** (280mg, 1eq), 2-(iodomethyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (0.32ml, 1.2eq) K_2CO_3(200mg, 1.2eq) was dissolved in the 20ml anhydrous DMF. And the mixture was stirred in R.T. for 12h. The mixture was quenched with water, extracted with EA, washed with brine, dried over Na_2SO_4. And the organic aqueous was removed under reduced pressure. The residue was purified by column chromatography to afford **1-9** (180 mg, 30.8% yield).

Synthesis of 7-((1H-pyrazol-1-yl)methyl)-2-hydroxy-2,3-dihydrobenzo[e][1,4,2]dioxaborinine-8-carboxylic acid (CB1)

The mixture of **1-9** (324 mg, 1.0 eq) in dioxane (4.0 mL) and 3N NaOH (4.0 mL) was stirred at r.t. for 2 hours. Then 4M HCl in dioxane was added to the above solution to adjust the pH to 1-2, stirred for 1 hours at room temperature. The

crude product was purified by reverse-phase prep-HPLC to afford **CB1** (15mg, 7% yield).

Scheme 2. The synthetic route for **CB2**



Synthesis of 5-bromo-4-methoxythiophene-3-carboxylic acid (2-2)

2-1 (900mg, 5.7 mmol) in 12 mL DCM was cooled to 0 °C. And Br_2 in DCM (3 mL) was slowly added to the above solution over 0.5h. The mixture was rotary evaporated to give the crude product **2-2**.

Synthesis of *tert*-butyl 5-bromo-4-methoxythiophene-3-carboxylate (2-3)

To the solution of crude product **2-2** (5.7 mmol, 1eq) in DCM (12 mL) was added N,N'-Dicyclohexylcarbodiimide (2.3g, 2 eq), DMAP (348 mg, 0.5 eq), *tert*-Butanol (17 ml). The resulting reaction mixture was stirred at room temperature overnight before it was concentrated to dryness. The residue was purified by column chromatography to afford compound **2-3** (1.6g, two-step reaction yield 96%)

Synthesis of *tert*-butyl 4-methoxy-5-vinylthiophene-3-carboxylate (2-4)

The **2-3** (1g,1eq), potassium vinyl trifluoroborate (547 mg, 1.2eq), $Pd(dppf)Cl_2$ (200 mg, 0.8eq) was dissolved in the isopropanol-water (21ml:7ml). The mixture was degassed with Ar for three times before adding Et_3N (0.71 ml, 1.5eq). The reaction mixture was stirred at 80℃ for 5h. Then the water was added to the solutions of reaction. Extract the aqueous layer with EA, wash with brine, dry over anhydrous Na_2SO_4. The residue was purified by column

chromatography to give the compound **2-4** (580mg, 71%yield).

Synthesis of *tert*-butyl 4-methoxy-5-(2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)ethyl)thiophene-3-carboxylate (2-5)

The **2-4** (580 mg, 1eq), PinBH (0.43ml, 1.2eq), $[IrCl(COD)]_2$ (128 mg, 0.08eq) and dppe (153 mg, 0.16eq) was dissolved in 20ml DCM. The solution was degassed three times and flushed with Ar. The mixture was stirred at 25℃ for 12h. Quench the reaction with methanol, the resulting residue was concentrated to dryness and was purified by column chromatography to afford the compound **2-5** (580mg, 65%yield).

Synthesis of 2-hydroxy-3,4-dihydro-2H-thieno[2,3-e][1,2]oxaborinine-7-carboxylic acid (CB2)

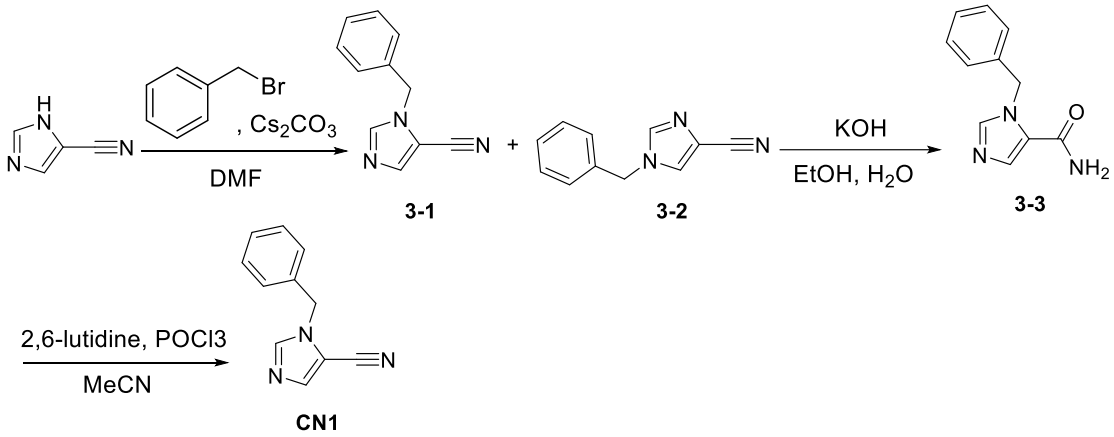
The soulution of **2-5** (293 mg, 1eq) in anhydrous CH_2Cl_2 was cooled to -78 °C. Then BCl_3 (1M in CH_2Cl_2, 4 mL, 5 equiv) was slowly added to the reaction mixture which was then stirred for 1 h at 0 °C. Then the reaction mixture was quenched by the addition of water. The aqueous phase was extracted with EA and the organic phase was concentrated in vacuo. The residue was recrystallized by using ethyl acetate and *n*-hexane to obtain **CB2** (40mg, 26%yield).

NMR and HPLC spectra of CB1 and CB2

The NMR and HPLC characterizations for compounds **CB1** and **CB2** are given in Supplementary Fig. 20-23.

1.3 Chemical synthesis and characterization of CN1 and CN2

Scheme 3. The synthetic route for **CN1**



Synthesis of 1-benzyl-1*H*-imidazole-5-carbonitrile and1-benzyl-1*H*-imidazole-4-carbonitrile (3-1 and 3-2)

The solution of benzyl bromide (1.88g, 11mmol) was added to dry DMF (40mL) with 1*H*-imidazole-4-carbonitrile (931mg, 10mmol) and cesium carbonate (3.91g, 12mmol) under argon. After the dropwise addition, the reaction was stirred for 16h at room temperature. Dilute the residue with EA and water, followed by extraction with EA three times. The combined organic phase was washed with brine, dried over anhydrous sodium sulfate and filtered, then concentrated under reduced pressure and the residue was purified by column chromatography to give mixture **3-1** and **3-2** (1.72g, 94% yield).

Synthesis of 1-benzyl-1H-imidazole-5-carboxamide (3-3)

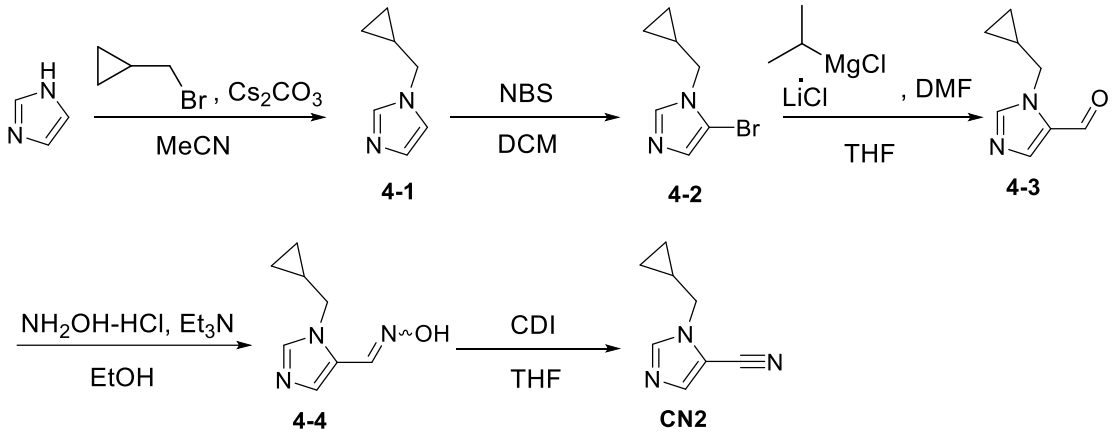
A mixed solution of water (8mL) and sodium hydroxide (1.28g, 32mmol) was added dropwise in ethanol dissolving mixture of **3-1** and **3-2**, then the reaction was refluxed for 2h at 65°C. After the reaction was complete as indicated by TLC, the mixture was diluted with water and extracted with EA three times. Dried over anhydrous sodium sulfate and filtered, the combined organic phase was concentrated under reduced pressure and the residue was purified by column chromatography to give compound **3-3** (338mg, 21% yield).

Synthesis of 1-benzyl-1H-imidazole-5-carbonitrile (CN1)

A mixture of phosphorus oxychloride (767mg, 5mmol), 2,6-lutidine (536mg,

5mmol) and dry MeCN (5mL) were cooled to 0°C under argon, then **3-3** (201mg, 1mmol) was added slowly. Remove cooling bath and the reaction was stirred for 16h at room temperature. The reaction was complete as indicated by TLC, then saturated sodium bicarbonate aqueous solution quenched the reaction solution. Water was added to the reaction mixture, followed by extraction with ethyl acetate three times and washed the combined organic phase with brine. Dry the extracts over anhydrous sodium sulfate and filter, then concentrated under reduced pressure. Purify the residue by column chromatography to obtain compound **CN1** (119mg, 65% yield).

Scheme 4. The synthetic route for **CN2**



Synthesis of 1-(cyclopropylmethyl)-1*H*-imidazole (4-1)

The solution of (bromomethyl)cyclopropane (2.23g, 16.5mmol) was added to dry MeCN (45mL) with imidazole (1.02g, 15mmol), cesium carbonate (5.86g, 18mmol) under argon, then the reaction was stirred for 16h at room temperature. Concentrate the reaction mixture in vacuo and the residue was purified by column chromatography to obtain **4-1** (1.79g, 98% yield).

Synthesis of 5-bromo-1-(cyclopropylmethyl)-1*H*-imidazole (4-2)

Cool the mixture of **4-1** (1.22g, 10mmol) and DCM (30mL) to 0°C, the solution of 1-Bromopyrrolidine-2,5-dione (1.96g, 11mmol) and DCM (100mL) was added dropwise to the mixture. Then the reaction was stirred for 4h at room temperature. After the reaction was complete as indicated by TLC, the reaction

was concentrated mixture in vacuo and purified by column chromatography to give **4-2** (683mg, 34% yield).

Synthesis of 1-(cyclopropylmethyl)-1*H*-imidazole-5-carbaldehyde (4-3)

Cool the solution of **4-2** (683mg, 3.4mmol) and THF (10.2mL) to 0°C, then Isopropylmagnesium chloride lithium chloride complex (6.54mL 1.3mol/L, 8.5mmol) was added dropwise to solution and stirred for 1h at room temperature under argon. After that, N,N-Dimethylformamide (746mg, 10.2mmol) was added slowly in mixture and stirred for 8h at room temperature. The reaction complete as indicated by TLC, saturated ammonium chloride aqueous solution quenched the reaction solution. Dilute the mixture with water and extract with EA three times. Dried over anhydrous sodium sulfate and filtered, the combined organic phase was concentrated under reduced pressure and purified by column chromatography to give **4-3** (386mg, 76% yield).

Synthesis of 1-(cyclopropylmethyl)-1H-imidazole-5-carbaldehyde oxime (4-4)

The mixture of **4-3** (386mg, 2.58mmol), hydroxylamine hydrochloride (269mg, 3.87mmol), triethylamine (574mg, 5.68mmol) and EtOH (11.6mL) were heated at 80°C for 3h. After the reaction complete, water was added to the reaction mixture, followed by extraction with ethyl acetate three times. Dry combined organic phase over anhydrous sodium sulfate and filter, then concentrate filtrate under reduced pressure and purify residue by column chromatography to give compound **4-4** (397mg, 93% yield).

Synthesis of 1-(cyclopropylmethyl)-1*H*-imidazole-5-carbonitrile (CN2)

Compound **4-4** (397mg, 2.40mmol) was added to dry THF (9.6mL) dissolving di(1*H*-imidazol-1-yl)methanone(1.17g, 7.2 mmol) and refluxed for 12h at 60°C under argon. After the reaction completed, concentrate the reaction mixture in vacuo and purify the residue by column chromatography to give **CN2**(278mg, 79%).

NMR and HPLC spectra of CN1 and CN2

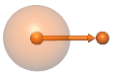
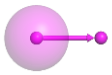
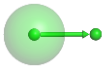
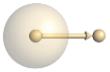
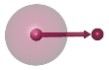
The NMR and HPLC characterizations for compounds **CN1** and **CN2** are given in Supplementary Fig. 24-29.

Section 2 Supplementary Tables

Supplementary Table 1 | The number of ligand-pharmacophore pairs in the LigPhore, CpxPhore and DockPhore datasets divided for training, validation, and testing.

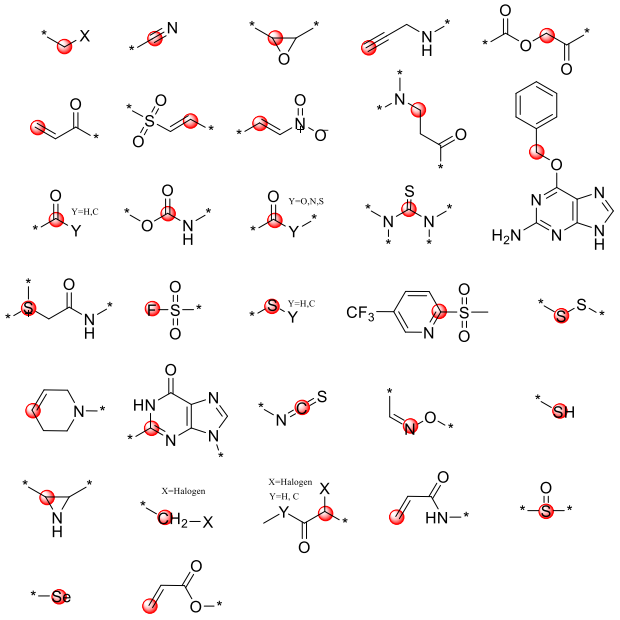
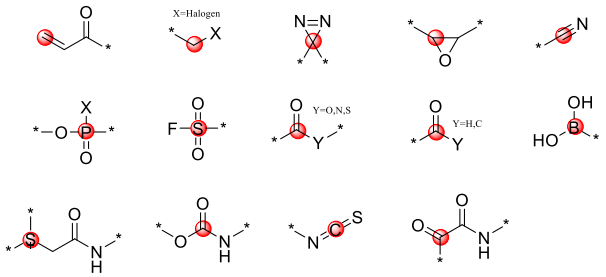
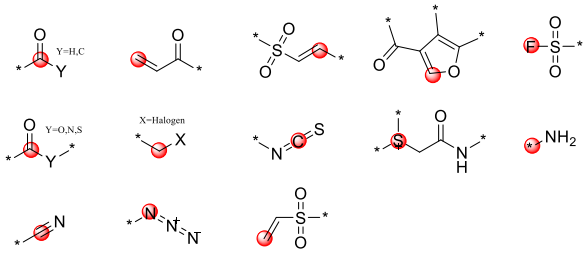
Dataset	Training set	Validation set	Test set
LigPhore	2,350,797	23,988	23,991
CpxPhore	12,581	735	269
DockPhore	76,203	8,468	96

Supplementary Table 2 | PyMOL display of pharmacophore features and their distributions in the training set.[a]

Feature Type	Display	Feature count		
		LigPhore training set	CpxPhore training set	DockPhore training set
Metal Coordination (MB)		1,838,768	2,989	0
Hydrogen Bond Donor (HD)		1,064,279	18,374	103,148
Hydrogen Bond Acceptor (HA)		4,149,336	25,845	169,668
Aromatic Ring (AR)		1,995,623	4,385	27,936
Halogen Bonding (XB)		337,608	296	942
Positively Charged Center (PO)		214,686	2,016	5
Negatively Charged Center (NE)		195,935	2,581	18,438
Hydrophobic (HY)		3,634,799	16,110	115,869
Covalent Bonding I (CV1)[b]		283,962	287	3
Covalent Bonding II (CV2)[c]		173,545	324	0
Covalent Bonding III (CV3)[d]		214,128	20	2
Covalent Bonding IV (CV4)[e]		2,872	5	0
Exclusion Sphere (EX)		236,836,194	984,826	5,798,989

[a] The detailed definitions of pharmacophore features can be found in Ref. 1. [b] CV1: a functional group capable of forming a covalent bond with cysteine residue (Cys-SH), such as electrophilic groups that target thiol groups in proteins. [c] CV2: a functional group capable of forming a covalent bond with tyrosine, serine, or threonine residue (Tyr/Ser/Thr-OH). [d] CV3: a functional group capable of forming a covalent bond with lysine residue (Lys-NH_2), typically targeting amino group for stable adduct formation. [e] CV4: a functional group capable of forming a covalent bond with glutamate or aspartate residue (Glu/Asp-COO-), often involving a carboxylate-targeting reactive group.

Supplementary Table 3 | The detailed definitions of the four covalent features (CV1, CV2, CV3, and CV4).

Covalent bonding type	Reaction residue	Covalent warhead
CV1	Cys-SH	
CV2	Tyr/Ser/Thr-OH	
CV3	Lys-NH2	
CV4	Glu/Asp-COO$^-$	

Supplementary Table 4 | The influence of bond diffusion, pretraining, and refinement training.

Model variant[a]	Validity (↑)	Mapping score (↑)	Docking score (↓)	PoseBusters (↑)	QED(↑)	SA (↑)	Diversity
PhoreGen	0.91	0.76	-8.03	0.74	0.61	0.71	0.65
NoPretraining	0.90	0.75	-7.85	0.47	0.61	0.69	0.66
NoRefineTraining	0.51	0.56	-7.28	0.66	0.52	0.67	0.64
NoBondDiffusion	0.89	0.67	-8.16	0.34	0.54	0.64	0.69

[a] The "NoPretraining" variant refers to the model being trained directly on the CpxPhore and DockPhore datasets, bypassing the pretraining phase on the LigPhore dataset. The "NoRefineTraining" variant refers to the model being trained solely on the LigPhore dataset, without refinement through the CpxPhore and DockPhore datasets. The "NoBondDiffusion" variant refers to the exclusion of chemical bond information in the diffusion-denoising process, with only atomic features being considered. Using the same set of 10 protein targets described in the main text of "**PhoreGen is highly practical for de novo drug design**", we generated 100 molecules per target for each variant and the complete PhoreGen model.

Supplementary Table 5 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 1/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6rnu_complex.phore	4	68	0.97	0.81	0.78	19
6npi_complex.phore	4	40	0.96	0.84	0.8	10
6i8m_complex.phore	3	87	0.95	0.53	0.69	29
6i8t_complex.phore	3	84	0.95	0.63	0.72	28
6npp_complex.phore	3	54	0.95	0.86	0.77	17
6hmt_complex.phore	11	69	0.94	0.78	0.57	36
6kqi_complex.phore	9	109	0.94	0.69	0.51	29
6jbe_complex.phore	4	51	0.94	0.88	0.73	12
6fe5_complex.phore	11	92	0.93	0.82	0.46	22
6hhp_complex.phore	11	68	0.93	0.76	0.57	36
6jb0_complex.phore	8	88	0.93	0.81	0.49	25
6jag_complex.phore	7	95	0.93	0.8	0.47	23
6jbb_complex.phore	7	90	0.93	0.82	0.52	24
6i61_complex.phore	6	70	0.93	0.81	0.62	18
6s57_complex.phore	5	72	0.93	0.78	0.58	24
6oio_complex.phore	4	82	0.93	0.74	0.72	27
6oir_complex.phore	4	73	0.93	0.81	0.74	25
6ffe_complex.phore	4	52	0.93	0.72	0.81	19
6nlj_complex.phore	4	47	0.93	0.33	0.46	17
6a1c_complex.phore	3	78	0.93	0.46	0.63	30
6k2n_complex.phore	3	49	0.93	0.73	0.78	21
6te6_complex.phore	10	93	0.92	0.76	0.49	32
6s55_complex.phore	8	66	0.92	0.77	0.52	25
6jb4_complex.phore	7	93	0.92	0.77	0.53	25
6jap_complex.phore	7	84	0.92	0.8	0.47	22
6oiq_complex.phore	4	90	0.92	0.64	0.72	28
6oip_complex.phore	4	83	0.92	0.6	0.71	29
6oin_complex.phore	4	79	0.92	0.75	0.75	24
6nri_complex.phore	3	71	0.92	0.46	0.76	27
6nrg_complex.phore	3	68	0.92	0.69	0.71	26
6jam_complex.phore	9	84	0.91	0.75	0.47	23
6jad_complex.phore	8	93	0.91	0.78	0.57	25
6j9y_complex.phore	8	88	0.91	0.77	0.54	22
6r7d_complex.phore	6	84	0.91	0.68	0.7	29
6cjp_complex.phore	5	84	0.91	0.63	0.71	30
6i75_complex.phore	4	61	0.91	0.68	0.73	26

Supplementary Table 6 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 2/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6pya_complex.phore	4	61	0.91	0.85	0.8	19
6qlp_complex.phore	3	66	0.91	0.56	0.75	26
6qlr_complex.phore	3	65	0.91	0.7	0.75	24
6qlu_complex.phore	3	64	0.91	0.66	0.74	26
6qlo_complex.phore	3	63	0.91	0.61	0.74	26
6qlq_complex.phore	3	61	0.91	0.58	0.74	27
6qln_complex.phore	3	60	0.91	0.61	0.76	24
6efk_complex.phore	15	82	0.9	0.2	0.51	41
6nsv_complex.phore	13	78	0.9	0.41	0.61	42
6jao_complex.phore	10	86	0.9	0.73	0.51	25
6qzh_complex.phore	10	85	0.9	0.66	0.6	33
6e6j_complex.phore	10	66	0.9	0.57	0.59	32
6j9w_complex.phore	9	83	0.9	0.7	0.47	23
6tel_complex.phore	8	107	0.9	0.73	0.52	34
6jan_complex.phore	8	86	0.9	0.82	0.53	23
6bqd_complex.phore	6	77	0.9	0.81	0.73	22
6ooy_complex.phore	6	75	0.9	0.61	0.66	24
6k3l_complex.phore	6	66	0.9	0.76	0.71	22
6ffg_complex.phore	6	52	0.9	0.74	0.72	25
6hoq_complex.phore	5	59	0.9	0.68	0.81	17
6jib_complex.phore	5	56	0.9	0.77	0.72	21
6kjf_complex.phore	4	94	0.9	0.66	0.68	30
6nrj_complex.phore	4	87	0.9	0.54	0.64	31
6pyb_complex.phore	3	75	0.9	0.67	0.74	25
6nxz_complex.phore	3	42	0.9	0.85	0.83	13
6ten_complex.phore	8	106	0.89	0.74	0.57	33
6fff_complex.phore	6	57	0.89	0.69	0.7	27
6rot_complex.phore	5	95	0.89	0.68	0.54	31
6hop_complex.phore	5	59	0.89	0.8	0.8	17
6py0_complex.phore	4	68	0.89	0.57	0.66	25
6h9v_complex.phore	4	30	0.89	0.61	0.81	15
6g3c_complex.phore	3	76	0.89	0.63	0.74	22
5zcu_complex.phore	3	74	0.89	0.63	0.68	23
6jon_complex.phore	9	73	0.88	0.57	0.68	27
6kjd_complex.phore	6	92	0.88	0.72	0.69	31
6n53_complex.phore	6	87	0.88	0.67	0.72	23
6sfc_complex.phore	6	77	0.88	0.69	0.64	21

Supplementary Table 7 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 3/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6cjs_complex.phore	6	71	0.88	0.68	0.7	30
6i64_complex.phore	5	70	0.88	0.74	0.56	15
6e4c_complex.phore	5	57	0.88	0.6	0.71	25
6qls_complex.phore	5	55	0.88	0.68	0.71	26
6jaq_complex.phore	5	54	0.88	0.88	0.61	13
6e3m_complex.phore	5	47	0.88	0.59	0.71	24
6i74_complex.phore	5	61	0.87	0.67	0.74	25
6i77_complex.phore	5	60	0.87	0.68	0.73	27
6i66_complex.phore	5	56	0.87	0.79	0.65	13
6i78_complex.phore	4	60	0.87	0.59	0.74	26
6cf7_complex.phore	3	72	0.87	0.28	0.65	36
6i76_complex.phore	3	65	0.87	0.62	0.73	28
6qlt_complex.phore	3	51	0.87	0.8	0.79	22
6inz_complex.phore	8	86	0.86	0.6	0.68	27
6i63_complex.phore	5	73	0.86	0.78	0.62	19
6i65_complex.phore	5	53	0.86	0.92	0.54	11
6n9l_complex.phore	11	81	0.85	0.54	0.49	27
6e13_complex.phore	5	66	0.85	0.71	0.75	22
6md6_complex.phore	5	51	0.85	0.59	0.7	22
6gdy_complex.phore	4	57	0.85	0.81	0.61	9
6i67_complex.phore	4	51	0.85	0.81	0.71	11
6qqq_complex.phore	3	82	0.85	0.84	0.74	22
6qre_complex.phore	3	68	0.85	0.82	0.74	16
5zr3_complex.phore	10	106	0.84	0.59	0.44	33
6i41_complex.phore	10	58	0.84	0.62	0.52	33
6g2f_complex.phore	7	67	0.84	0.68	0.66	22
5zk7_complex.phore	5	47	0.84	0.35	0.74	36
6qrg_complex.phore	4	88	0.84	0.72	0.74	25
6i7a_complex.phore	9	58	0.83	0.5	0.57	39
6oxz_complex.phore	7	117	0.83	0.18	0.72	41
6g2c_complex.phore	7	61	0.83	0.69	0.69	19
6gbw_complex.phore	6	80	0.83	0.72	0.56	32
6oxq_complex.phore	5	118	0.83	0.37	0.67	40
6qfe_complex.phore	4	71	0.83	0.58	0.73	24
6s56_complex.phore	3	71	0.83	0.78	0.76	24
6uvy_complex.phore	3	66	0.83	1	0.9	14
6op0_complex.phore	10	98	0.82	0.43	0.6	33

Supplementary Table 8 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 4/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6i68_complex.phore	10	57	0.82	0.47	0.56	41
6k05_complex.phore	9	63	0.82	0.73	0.48	30
6ooz_complex.phore	8	79	0.82	0.62	0.64	26
6oxs_complex.phore	6	117	0.82	0.13	0.72	44
6oxt_complex.phore	6	115	0.82	0.12	0.67	44
6g2e_complex.phore	6	59	0.82	0.6	0.79	19
6a87_complex.phore	6	40	0.82	0.44	0.69	22
6oxw_complex.phore	5	127	0.82	0.33	0.68	43
6uwp_complex.phore	5	84	0.82	0.65	0.71	32
6qr2_complex.phore	5	80	0.82	0.59	0.72	23
6jsf_complex.phore	4	93	0.82	0.57	0.69	31
6jsg_complex.phore	4	79	0.82	0.6	0.71	27
6qqu_complex.phore	3	77	0.82	0.61	0.74	18
6qqw_complex.phore	3	75	0.82	0.57	0.76	20
6jt3_complex.phore	7	85	0.81	0.61	0.68	30
6oxu_complex.phore	6	113	0.81	0.25	0.71	43
6g2b_complex.phore	6	59	0.81	0.91	0.8	15
6jse_complex.phore	5	107	0.81	0.55	0.69	34
6jsz_complex.phore	5	81	0.81	0.46	0.66	33
6n4e_complex.phore	5	80	0.81	0.55	0.62	30
6jid_complex.phore	5	61	0.81	0.55	0.73	22
6qr0_complex.phore	4	84	0.81	0.54	0.73	21
6e6v_complex.phore	3	33	0.81	0.75	0.85	17
6i5p_complex.phore	9	51	0.8	0.47	0.61	37
6qr4_complex.phore	3	88	0.8	0.59	0.72	23
6qrf_complex.phore	3	64	0.8	0.87	0.79	14
6rtn_complex.phore	7	62	0.79	0.8	0.71	21
6n96_complex.phore	7	40	0.79	0.45	0.79	21
6jsn_complex.phore	6	88	0.79	0.65	0.67	33
6h7d_complex.phore	6	45	0.79	0.74	0.69	12
6qr9_complex.phore	4	92	0.79	0.46	0.72	27
6qqz_complex.phore	4	81	0.79	0.68	0.73	21
6quv_complex.phore	3	54	0.79	0.63	0.75	27
6nv7_complex.phore	11	120	0.78	0.61	0.6	35
6n93_complex.phore	9	79	0.78	0.28	0.7	41
6oxr_complex.phore	8	119	0.78	0	0.53	47
6qtw_complex.phore	8	96	0.78	0.54	0.57	46

Supplementary Table 9 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 5/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6oy1_complex.phore	7	124	0.78	0.18	0.7	42
6oy2_complex.phore	6	126	0.78	0.36	0.68	44
6oxy_complex.phore	6	124	0.78	0.24	0.66	44
6qr1_complex.phore	3	96	0.78	0.58	0.68	25
6od6_complex.phore	3	86	0.78	0.33	0.66	28
6moa_complex.phore	9	58	0.77	0.31	0.51	32
6oxp_complex.phore	6	110	0.77	0.16	0.63	40
6qtq_complex.phore	4	87	0.77	0.72	0.65	37
6n97_complex.phore	9	106	0.76	0.25	0.67	43
6oy0_complex.phore	6	125	0.76	0.31	0.65	44
6qgf_complex.phore	5	73	0.76	0.55	0.73	30
6a73_complex.phore	10	45	0.75	0.13	0.71	42
6n92_complex.phore	9	106	0.75	0.22	0.64	48
6n94_complex.phore	7	101	0.75	0.2	0.68	43
6p8z_complex.phore	6	91	0.75	0.46	0.58	36
6qmr_complex.phore	6	83	0.75	0.61	0.73	27
6oi8_complex.phore	6	76	0.75	1		25
6nv9_complex.phore	10	131	0.74	0.34	0.62	39
6oxv_complex.phore	6	124	0.74	0.29	0.68	42
6k1s_complex.phore	4	95	0.74	0.5	0.38	28
6jut_complex.phore	7	113	0.73	0.29	0.52	37
6qge_complex.phore	5	70	0.73	0.51	0.74	29
6qts_complex.phore	4	96	0.73	0.56	0.66	41
5zlf_complex.phore	4	69	0.73	0.52	0.74	29
6uwv_complex.phore	3	89	0.73	0		35
6st3_complex.phore	4	77	0.72	0.64	0.77	28
6np4_complex.phore	8	87	0.71	0.5	0.61	38
6ftf_complex.phore	6	77	0.71	0.7	0.65	17
6n0m_complex.phore	6	59	0.71	0.8	0.45	11
6qmt_complex.phore	5	77	0.71	0.53	0.74	25
6g25_complex.phore	5	60	0.71	0.67	0.79	18
6hhr_complex.phore	3	71	0.71	0.49	0.82	21
6n55_complex.phore	3	46	0.71	0.39	0.73	16
6np5_complex.phore	10	78	0.7	0	0.68	40
6hhi_complex.phore	6	123	0.7	0.41	0.56	41
6os6_complex.phore	5	68	0.7	0.85	0.75	15
6os5_complex.phore	5	66	0.7	0.78	0.77	15

Supplementary Table 10 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 6/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6hv2_complex.phore	11	74	0.69	0.66	0.46	30
6nw3_complex.phore	9	131	0.69	0.59	0.64	41
6g2o_complex.phore	8	69	0.69	0.49	0.65	30
6s9x_complex.phore	5	147	0.69	0.52	0.51	49
6e3z_complex.phore	4	84	0.69	0.43	0.75	32
6hhh_complex.phore	5	134	0.68	0.52	0.54	43
6e5s_complex.phore	6	90	0.67	0.7	0.73	24
6pgo_complex.phore	4	82	0.67	0.55	0.75	32
6qtx_complex.phore	3	94	0.67	0.64	0.71	36
6p8x_complex.phore	6	100	0.66	0.31	0.73	33
6e7m_complex.phore	6	87	0.66	0.69	0.71	23
6qrd_complex.phore	3	98	0.66	0.43	0.72	31
6mhd_complex.phore	5	58	0.65	0.47	0.71	26
6uvv_complex.phore	3	58	0.65	1	0.42	13
6mo9_complex.phore	3	23	0.65	0.41	0.79	12
6qw8_complex.phore	8	62	0.64	0.73	0.75	23
6op9_complex.phore	3	93	0.64	0.59	0.75	30
6np3_complex.phore	10	86	0.63	0	0.66	40
6jmf_complex.phore	5	68	0.63	0.58	0.72	23
6gwe_complex.phore	5	99	0.62	0.43	0.71	34
6pz4_complex.phore	4	96	0.62	0.43	0.77	35
6p8y_complex.phore	4	91	0.62	0.33	0.52	43
6pgp_complex.phore	3	106	0.62	0.57	0.6	38
6mo8_complex.phore	7	52	0.61	0.06	0.66	25
6oim_complex.phore	3	103	0.61	0.53	0.6	40
6o5u_complex.phore	11	78	0.6	0	0.71	41
6np2_complex.phore	10	88	0.6	0.16	0.64	41
6i9a_complex.phore	7	77	0.57	0.53	0.6	36
6gj8_complex.phore	4	57	0.57	0.57	0.74	28
6qr7_complex.phore	5	96	0.56	0.68	0.67	29
6e3p_complex.phore	4	59	0.56	0	0.7	26
6qr3_complex.phore	4	79	0.55	0.58	0.7	24
6mhc_complex.phore	3	47	0.55	0.72	0.82	19
6mo2_complex.phore	6	66	0.53	0.27	0.79	25
6hhg_complex.phore	4	130	0.53	0.39	0.54	47
6g9f_complex.phore	3	53	0.53	0.77	0.81	13
6e6w_complex.phore	5	54	0.52	0		31

Supplementary Table 11 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 7/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6mo0_complex.phore	4	68	0.52	0.45	0.81	28
6quw_complex.phore	3	43	0.52	0.76	0.83	14
5zxk_complex.phore	6	53	0.51	0.64	0.64	12
6cjr_complex.phore	6	98	0.5	0		36
6e3o_complex.phore	5	39	0.49	0.41	0.8	22
6ufo_complex.phore	6	69	0.48	0.66	0.76	19
6r4k_complex.phore	5	69	0.46	0.25	0.7	29
6ckl_complex.phore	5	81	0.45	0	0.9	26
6qwi_complex.phore	5	65	0.45	0.22	0.52	33
6izq_complex.phore	4	61	0.45	0.96	0.6	14
6iql_complex.phore	3	74	0.45	0.71	0.69	26
6g27_complex.phore	5	59	0.44	0.82	0.77	19
6ufn_complex.phore	3	70	0.44	0.63	0.82	20
6g29_complex.phore	4	60	0.43	0.69	0.75	20
6cyh_complex.phore	3	71	0.43	0.12	0.66	25
6e3n_complex.phore	5	40	0.42	0		11
6hhj_complex.phore	4	135	0.42	0.33	0.53	49
6k04_complex.phore	9	52	0.41	0	0.71	30
6uhu_complex.phore	5	70	0.41	0.71	0.74	19
6h14_complex.phore	11	130	0.39	0.12	0.66	39
6j06_complex.phore	8	57	0.35	0.19	0.61	21
6uhv_complex.phore	5	70	0.29	0.49	0.73	20
6nt2_complex.phore	3	86	0.29	0.56	0.79	31
6uii_complex.phore	3	62	0.19	0.72	0.77	20
6ott_complex.phore	10	84	0.15	0.31	0.77	31
6h12_complex.phore	9	123	0.1	0.14	0.73	38
6ibz_complex.phore	5	102	0.1	0.2	0.7	32
6uil_complex.phore	6	83	0.07	0	0.66	26
6o0h_complex.phore	6	101	0.04	0.12	0.49	31
6ibx_complex.phore	5	101	0.01	0	0.67	32
6ic0_complex.phore	4	104	-0.01	0.06	0.59	34
6iby_complex.phore	5	98	-0.04	0	0.48	33
6hzb_complex.phore	11	110	-	-	-	-
6hza_complex.phore	10	114	-	-	-	-
6h13_complex.phore	8	136	-	-	-	-
6oxx_complex.phore	8	115	-	-	-	-
6a6k_complex.phore	7	95	-	-	-	-

Supplementary Table 12 | The distributions of features and exclusion spheres in the pharmacophore models of the CpxPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 8/8).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
6cjj_complex.phore	7	72	-	-	-	-
6rz6_complex.phore	6	146	-	-	-	-
6n4b_complex.phore	6	101	-	-	-	-
6mo7_complex.phore	6	66	-	-	-	-
6o3x_complex.phore	6	59	-	-	-	-
6gzy_complex.phore	5	58	-	-	-	-
6uim_complex.phore	4	67	-	-	-	-
6cyg_complex.phore	4	64	-	-	-	-
6dz3_complex.phore	3	94	-	-	-	-
6qxd_complex.phore	3	79	-	-	-	-
6uvp_complex.phore	3	76	-	-	-	-

Supplementary Table 13 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 1/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
71285302_1_NE.phore	6	74	0.92	0.55	0.67	34
Bsp2079270_0_Bsp2.phore	7	140	0.9	0.69	0.59	26
529610175_0_primary.phore	5	125	0.9	0.72	0.61	25
1875333721_1_primary.phore	4	127	0.89	0.84	0.68	22
426491591_1_NE.phore	7	113	0.89	0.71	0.7	26
Bsp3033181_2_Bsp3.phore	8	148	0.88	0.68	0.55	31
35190379_2_PO.phore	8	140	0.88	0.48	0.61	31
29264599_0_primary.phore	7	108	0.88	0.55	0.69	25
84136818_1_primary.phore	6	72	0.88	0.77	0.7	21
65382160_0_primary.phore	6	70	0.88	0.78	0.75	18
11573541_2_XB.phore	6	132	0.87	0.64	0.54	28
103345949_1_XB.phore	5	111	0.87	0.55	0.67	26
193587068_0_primary.phore	6	94	0.87	0.73	0.69	24
1688811_0_primary.phore	5	61	0.87	0.8	0.77	18
3861007_1_primary.phore	3	20	0.87	0.4	0.78	13
96399391_0_NE.phore	9	164	0.86	0.47	0.45	36
17169926_2_XB.phore	7	157	0.86	0.54	0.59	28
95524945_0_primary.phore	7	143	0.86	0.53	0.65	28
454290358_2_primary.phore	7	84	0.86	0.67	0.59	24
1875354387_1_NE.phore	4	77	0.86	0.76	0.75	18
95957685_1_primary.phore	5	70	0.86	0.64	0.82	17
2663964_1_XB.phore	10	154	0.85	0.45	0.49	38
409224318_1_primary.phore	7	133	0.85	0.35	0.54	34
67515028_1_primary.phore	6	122	0.85	0.63	0.65	27
2421444_2_primary.phore	6	106	0.85	0.71	0.67	24
604405975_1_XB.phore	6	105	0.85	0.46	0.5	34
100745665_1_primary.phore	4	92	0.85	0.83	0.7	24
180866697_0_primary.phore	6	91	0.85	0.63	0.69	26
84207944_1_primary.phore	6	87	0.85	0.62	0.63	27
25814958_1_primary.phore	8	144	0.84	0.68	0.51	32
811548595_1_primary.phore	6	138	0.84	0.7	0.6	29
12549967_2_primary.phore	9	130	0.84	0.58	0.61	32
Bsp2050748_2_Bsp2.phore	7	91	0.84	0.67	0.66	26
91364793_2_primary.phore	6	85	0.84	0.68	0.72	23
4992251_2_NE.phore	5	72	0.84	0.84	0.76	15
79048749_2_primary.phore	6	115	0.83	0.61	0.72	27
329977350_2_primary.phore	4	102	0.83	0.8	0.76	21

Supplementary Table 14 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 2/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
96184862_0_primary.phore	5	96	0.83	0.7	0.77	22
524730445_0_CV.phore	6	87	0.83	0.62	0.58	28
12418049_0_primary.phore	6	74	0.83	0.58	0.72	24
2325804708_0_primary.phore	5	67	0.83	0.7	0.75	20
84705848_1_primary.phore	5	61	0.83	0.7	0.65	29
1875354664_2_primary.phore	7	166	0.82	0.69	0.56	33
56134325_2_primary.phore	4	120	0.82	0.66	0.75	23
71908712_1_CV.phore	6	108	0.82	0.77	0.67	24
585116406_2_primary.phore	5	103	0.82	0.73	0.73	22
585656336_0_primary.phore	5	100	0.82	0.62	0.69	20
32908968_1_primary.phore	5	98	0.82	0.78	0.8	15
9252564_0_XB.phore	9	151	0.81	0.58	0.4	33
91663153_1_NE.phore	7	137	0.81	0.69	0.57	29
12558126_1_primary.phore	5	123	0.81	0.77	0.74	24
12219593_2_primary.phore	7	108	0.81	0.73	0.69	23
331484437_1_XB.phore	6	107	0.81	0.62	0.58	28
39929403_1_primary.phore	5	101	0.81	0.7	0.67	27
257209923_0_XB.phore	6	101	0.81	0.71	0.63	25
96495310_0_XB.phore	5	99	0.81	0.8	0.79	18
1875406970_2_primary.phore	5	89	0.81	0.69	0.72	22
19857809_2_CV.phore	8	129	0.8	0.6	0.5	37
67803787_2_NE.phore	7	120	0.8	0.72	0.68	25
299797142_0_primary.phore	7	120	0.8	0.47	0.62	27
58305207_1_primary.phore	6	117	0.8	0.63	0.74	26
14093848_0_primary.phore	6	115	0.8	0.73	0.64	24
Bsp3034845_1_Bsp3.phore	7	112	0.8	0.76	0.69	26
96833603_2_NE.phore	6	107	0.8	0.71	0.66	25
2325850999_2_primary.phore	8	94	0.8	0.51	0.5	39
67673026_2_NE.phore	6	89	0.8	0.46	0.75	20
1515968226_0_primary.phore	6	83	0.8	0.82	0.74	20
72240075_2_primary.phore	6	76	0.8	-	-	-
67846728_0_primary.phore	5	75	0.8	-	-	-
843240365_0_primary.phore	6	73	0.8	0.69	0.73	21
1652005506_1_primary.phore	6	73	0.8	0.75	0.68	22
824630600_0_primary.phore	5	151	0.79	0.63	0.64	32
38339194_0_primary.phore	9	150	0.79	0.47	0.57	33
40104528_1_primary.phore	7	138	0.79	0.72	0.62	33

Supplementary Table 15 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 3/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
125617807_2_primary.phore	6	121	0.79	0.74	0.73	22
225423483_2_primary.phore	6	121	0.79	0.68	0.69	30
643588156_2_XB.phore	9	117	0.79	0.61	0.54	32
89738451_1_primary.phore	6	111	0.79	0.73	0.67	24
1775985243_0_primary.phore	5	110	0.79	0.46	0.72	28
49543599_1_primary.phore	5	107	0.79	0.69	0.63	21
330064762_1_primary.phore	5	106	0.79	0.63	0.68	24
306139283_1_primary.phore	7	94	0.79	0.47	0.7	26
626751656_2_NE.phore	7	87	0.79	0.54	0.6	34
11873567_2_primary.phore	5	83	0.79	0.59	0.67	24
1875263761_1_primary.phore	5	69	0.79	0.72	0.82	18
Bsp2059580_1_Bsp2.phore	8	62	0.79	0.44	0.77	27
225327709_1_primary.phore	6	160	0.78	0.48	0.62	30
20272304_0_primary.phore	10	124	0.78	0.72	0.56	33
59520325_0_XB.phore	7	122	0.78	0.64	0.67	25
Bsp2062734_2_Bsp2.phore	7	113	0.78	0.65	0.71	27
55358617_2_primary.phore	6	112	0.78	0.77	0.65	23
67849959_2_NE.phore	6	106	0.78	0.59	0.7	21
820708439_2_primary.phore	5	86	0.78	0.41	0.66	27
72275418_2_primary.phore	5	86	0.78	0.69	0.63	26
Bsp2044997_0_Bsp2.phore	8	81	0.78	0.47	0.64	29
12321054_0_CV.phore	7	78	0.78	0.63	0.66	27
299781879_0_primary.phore	5	77	0.78	0.51	0.76	24
35480748_0_primary.phore	7	70	0.78	0.69	0.74	20
72310516_0_primary.phore	6	53	0.78	0.63	0.72	18
69351711_0_primary.phore	5	140	0.77	0.78	0.75	20
Bsp2054417_1_Bsp2.phore	9	134	0.77	0.38	0.62	32
2325836711_2_primary.phore	10	131	0.77	0.48	0.5	38
72436197_0_NE.phore	6	128	0.77	0.5	0.64	27
438087621_1_primary.phore	5	122	0.77	0.79	0.64	26
14713180_2_primary.phore	5	121	0.77	0.65	0.75	25
597936073_2_primary.phore	6	119	0.77	0.52	0.68	25
1560014_0_primary.phore	6	111	0.77	0.61	0.59	25
1565527945_0_NE.phore	5	109	0.77	0.54	0.72	23
91845982_0_primary.phore	4	89	0.77	0.72	0.73	23
1065645603_0_primary.phore	4	85	0.77	0.73	0.79	20
1857704360_0_NE.phore	7	75	0.77	0.72	0.72	22

Supplementary Table 16 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 4/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
17126175_2_primary.phore	6	55	0.77	0.56	0.65	23
9529620_0_primary.phore	9	135	0.76	0.36	0.56	37
Bsp3014803_1_Bsp3.phore	7	124	0.76	0.53	0.62	31
95479399_1_primary.phore	7	122	0.76	0.7	0.64	26
91848242_0_NE.phore	4	115	0.76	0.6	0.68	27
1857671358_0_NE.phore	5	113	0.76	0.73	0.65	25
96183804_0_primary.phore	5	106	0.76	0.68	0.65	29
80648250_0_PO.phore	6	105	0.76	0.79	0.69	23
64897356_0_primary.phore	6	105	0.76	-	-	-
666034661_0_primary.phore	7	93	0.76	0.68	0.68	29
95350207_2_primary.phore	5	90	0.76	0.82	0.79	16
2056225271_2_XB.phore	6	87	0.76	0.58	0.7	28
645146702_0_primary.phore	4	75	0.76	0.76	0.81	16
2055759311_1_primary.phore	6	66	0.76	0.62	0.72	25
952880981_2_primary.phore	4	61	0.76	0.7	0.79	19
952981638_2_primary.phore	6	163	0.75	0.42	0.6	32
100834862_2_PO.phore	9	153	0.75	0.56	0.55	38
1775984136_1_NE.phore	6	123	0.75	0.51	0.7	29
1875284386_2_primary.phore	5	116	0.75	0.76	0.71	19
426433652_0_primary.phore	5	95	0.75	0.5	0.62	28
13756233_0_primary.phore	4	85	0.75	0.68	0.82	16
659404557_1_primary.phore	5	83	0.75	0.66	0.6	33
9683247_1_CV.phore	5	75	0.75	0.81	0.7	20
Bsp3006608_2_Bsp3.phore	6	70	0.75	0.6	0.71	26
665275888_0_primary.phore	6	53	0.75	0.66	0.78	19
32932962_1_primary.phore	7	146	0.74	0.71	0.62	27
6782458_0_XB.phore	6	129	0.74	0.53	0.63	30
76068376_1_primary.phore	6	124	0.74	0.66	0.67	23
1875360041_0_primary.phore	7	120	0.74	0.49	0.59	35
2768955_0_primary.phore	8	117	0.74	0.49	0.64	30
72403351_2_XB.phore	6	114	0.74	0.57	0.67	24
25078158_0_primary.phore	7	109	0.74	0.52	0.67	26
96246549_0_CV.phore	8	106	0.74	0.64	0.54	26
71903083_2_primary.phore	5	100	0.74	0.67	0.64	23
77425916_2_primary.phore	6	98	0.74	0.8	0.77	19
3525851_0_CV.phore	6	89	0.74	0.59	0.65	28
95358728_1_primary.phore	4	88	0.74	0.62	0.65	29

Supplementary Table 17 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 5/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
97334196_0_primary.phore	5	80	0.74	0.64	0.79	18
215579617_1_CV.phore	5	72	0.74	0.55	0.71	24
656661034_0_primary.phore	6	70	0.74	0.68	0.71	23
9413422_1_primary.phore	4	151	0.73	0.56	0.64	30
Bsp3018660_1_Bsp3.phore	9	122	0.73	0.57	0.6	30
1775994271_1_primary.phore	6	119	0.73	0.52	0.69	24
96113733_1_CV.phore	5	118	0.73	0.57	0.6	29
Bsp3024782_0_Bsp3.phore	9	111	0.73	0.46	0.68	31
839071650_2_primary.phore	7	108	0.73	0.5	0.63	26
17198016_2_primary.phore	7	106	0.73	0.68	0.58	28
1525613350_0_primary.phore	5	91	0.73	0.61	0.73	22
328636010_2_CV.phore	7	84	0.73	0.61	0.69	22
97742792_2_primary.phore	4	78	0.73	0.79	0.73	22
2325800058_2_primary.phore	4	64	0.73	0.72	0.72	19
22242950_1_primary.phore	8	161	0.72	0.51	0.63	33
2325757123_1_primary.phore	7	150	0.72	0.46	0.59	33
1848353655_2_NE.phore	7	124	0.72	0.71	0.69	25
36708374_2_primary.phore	4	123	0.72	0.71	0.71	22
2325809253_2_primary.phore	6	119	0.72	0.47	0.65	24
100837613_1_primary.phore	5	107	0.72	0.61	0.71	23
95382278_1_NE.phore	7	105	0.72	0.6	0.65	32
290619333_1_CV.phore	7	99	0.72	0.56	0.71	29
95443533_0_primary.phore	5	96	0.72	0.65	0.72	21
Bsp2044869_0_Bsp2.phore	6	96	0.72	0.62	0.69	29
252491700_1_primary.phore	6	93	0.72	0.6	0.78	22
931321167_1_primary.phore	6	88	0.72	0.65	0.79	18
1565525840_1_NE.phore	5	85	0.72	0.63	0.75	23
Bsp2054272_2_Bsp2.phore	7	74	0.72	0.48	0.71	26
823836880_2_primary.phore	5	66	0.72	0.7	0.82	17
100762197_0_primary.phore	6	60	0.72	0.75	0.77	21
584890624_1_primary.phore	6	132	0.71	0.67	0.65	23
Bsp2061911_1_Bsp2.phore	10	128	0.71	0.56	0.65	29
65526943_1_primary.phore	5	126	0.71	0.64	0.74	21
1875392561_0_NE.phore	6	124	0.71	0.56	0.66	23
71770906_2_primary.phore	6	120	0.71	0.77	0.67	24
881314929_2_primary.phore	4	119	0.71	0.71	0.71	24
58324482_0_primary.phore	8	114	0.71	0.51	0.61	33

Supplementary Table 18 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 6/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
2332238269_1_primary.phore	7	103	0.71	0.55	0.6	32
189432232_2_primary.phore	7	95	0.71	0.63	0.69	23
349952066_2_primary.phore	5	66	0.71	0.7	0.81	20
3139132_1_primary.phore	7	160	0.7	0.38	0.66	29
2325858971_1_primary.phore	7	158	0.7	0.59	0.66	29
Bsp3020465_0_Bsp3.phore	9	139	0.7	0.44	0.55	36
1875312525_0_primary.phore	6	118	0.7	0.69	0.69	25
Bsp3027911_1_Bsp3.phore	8	110	0.7	0.63	0.63	29
9884648_2_CV.phore	9	108	0.7	0.46	0.64	30
1875268062_2_primary.phore	5	102	0.7	0.63	0.71	23
96254005_2_primary.phore	5	97	0.7	0.46	0.77	23
19312784_2_PO.phore	5	95	0.7	0.75	0.64	21
45070122_0_primary.phore	4	94	0.7	0.71	0.71	18
91364616_2_primary.phore	6	76	0.7	0.62	0.72	25
66827109_2_XB.phore	6	74	0.7	0.71	0.78	22
19576739_0_primary.phore	6	74	0.7	-	-	-
1506415359_2_primary.phore	4	46	0.7	0.61	0.83	17
32541217_2_primary.phore	5	114	0.69	0.79	0.74	22
89601368_1_primary.phore	6	105	0.69	0.73	0.7	22
1565546300_1_primary.phore	7	87	0.69	0.61	0.6	27
Bsp3015341_2_Bsp3.phore	7	80	0.69	0.45	0.6	34
426471079_2_primary.phore	6	77	0.69	0.66	0.73	22
238701781_1_NE.phore	3	56	0.69	0.76	0.84	12
1615038130_1_primary.phore	6	131	0.68	0.62	0.64	24
1615552295_0_primary.phore	6	122	0.68	0.67	0.66	25
252501682_2_primary.phore	6	117	0.68	0.55	0.72	26
2325820648_1_primary.phore	5	113	0.68	0.56	0.75	23
26082502_0_PO.phore	6	112	0.68	0.51	0.69	22
2968243_1_PO.phore	7	89	0.68	0.83	0.57	26
36353294_2_CV.phore	5	83	0.68	0.57	0.76	25
575630220_0_primary.phore	6	77	0.68	0.48	0.62	33
83791600_1_XB.phore	5	68	0.68	0.76	0.7	19
15480679_1_XB.phore	6	67	0.68	0.56	0.7	29
65604853_0_primary.phore	4	54	0.68	0.71	0.85	16
15003946_1_primary.phore	9	133	0.67	0.63	0.59	29
1775990215_1_primary.phore	6	122	0.67	0.64	0.68	25
624940_2_primary.phore	7	117	0.67	0.57	0.6	27

Supplementary Table 19 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 7/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
95480394_2_CV.phore	8	107	0.67	0.42	0.64	27
96491856_0_CV.phore	5	100	0.67	0.68	0.7	24
567915914_0_PO.phore	7	99	0.67	0.78	0.66	25
218966607_2_primary.phore	6	64	0.67	0.53	0.7	26
112244137_1_PO.phore	11	192	0.66	0.36	0.49	41
28853060_0_primary.phore	7	136	0.66	0.52	0.53	33
843219319_0_primary.phore	5	121	0.66	0.56	0.68	20
593910126_2_primary.phore	6	115	0.66	0.71	0.65	26
8765440_1_CV.phore	8	110	0.66	0.47	0.62	31
14608591_1_primary.phore	6	98	0.66	0.67	0.7	25
14167894_0_primary.phore	4	84	0.66	0.76	0.81	19
2325790323_0_primary.phore	5	69	0.66	0.67	0.78	18
25338092_2_PO.phore	6	64	0.66	0.61	0.74	24
241586459_0_primary.phore	6	56	0.66	0.62	0.8	16
5386042_2_primary.phore	10	124	0.65	0.26	0.58	26
49441843_0_PO.phore	5	120	0.65	0.71	0.78	20
670453394_2_primary.phore	5	103	0.65	0.63	0.71	23
684875534_2_primary.phore	5	99	0.65	0.58	0.74	20
89799886_0_PO.phore	5	73	0.65	0.82	0.78	16
548455638_0_primary.phore	6	69	0.65	0.47	0.73	24
585123589_1_primary.phore	5	67	0.65	0.71	0.8	20
1704311235_0_primary.phore	7	128	0.64	0.61	0.59	29
27349385_2_CV.phore	5	121	0.64	0.59	0.65	29
10499000_0_CV.phore	6	121	0.64	0.71	0.6	29
828395482_1_primary.phore	6	120	0.64	0.59	0.71	28
8937824_1_XB.phore	5	101	0.64	0.58	0.78	19
Bsp2075951_1_Bsp2.phore	6	95	0.64	0.5	0.74	26
3168016_2_XB.phore	6	91	0.64	0.62	0.69	23
387366_2_primary.phore	4	80	0.64	0.58	0.78	19
9704292_0_PO.phore	8	73	0.64	0.63	0.63	27
Bsp3033089_0_Bsp3.phore	6	72	0.64	0.57	0.76	21
5257532_1_PO.phore	6	66	0.64	0.56	0.78	20
26896217_1_primary.phore	5	35	0.64	0.69	0.84	13
6012125_0_primary.phore	7	145	0.63	0.59	0.58	34
63199726_0_XB.phore	7	125	0.63	0.67	0.66	28
9513545_1_primary.phore	8	120	0.63	0.39	0.64	28
14897455_2_PO.phore	5	99	0.63	0.58	0.72	24

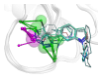
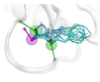
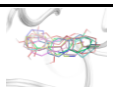
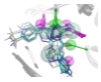
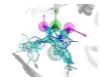
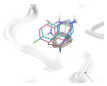
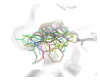
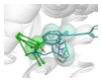
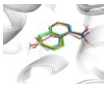
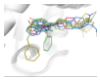
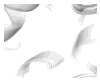
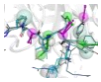
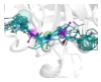
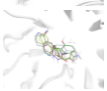
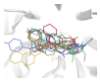
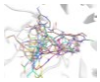
Supplementary Table 20 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 8/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
19626863_1_primary.phore	5	97	0.63	0.64	0.74	21
89984912_1_primary.phore	8	86	0.63	0.55	0.68	22
Bsp3024255_0_Bsp3.phore	6	76	0.63	0.61	0.72	28
12324027_2_primary.phore	6	145	0.62	0.44	0.61	24
72419051_0_CV.phore	7	113	0.62	0.42	0.64	26
7980101_2_primary.phore	6	112	0.62	0.63	0.7	25
97155761_2_primary.phore	6	83	0.62	0.66	0.76	20
35977088_2_primary.phore	3	68	0.62	0.8	0.87	12
2325893109_1_primary.phore	5	62	0.62	0.69	0.82	15
1506405933_1_primary.phore	5	56	0.62	0.67	0.73	20
13624089_0_CV.phore	5	145	0.61	0.6	0.65	29
13434330_1_PO.phore	6	133	0.61	0.69	0.52	26
211138084_2_CV.phore	5	108	0.61	0.65	0.55	22
1506420921_0_primary.phore	4	72	0.61	0.75	0.75	15
178757042_1_CV.phore	5	143	0.6	0.65	0.57	28
71856774_2_primary.phore	4	141	0.6	0.72	0.65	26
41135073_0_XB.phore	6	87	0.6	0.71	0.72	20
2325865508_1_primary.phore	4	76	0.6	0.5	0.69	23
827969666_0_primary.phore	5	55	0.6	0.4	0.68	27
25558887_1_XB.phore	8	115	0.59	0.4	0.53	32
72271936_1_PO.phore	4	96	0.58	0.78	0.74	22
409380061_2_PO.phore	6	90	0.58	0.47	0.71	25
Bsp2062290_2_Bsp2.phore	5	52	0.57	0.68	0.82	12
78876109_0_PO.phore	6	97	0.56	0.67	0.72	20
225583408_2_CV.phore	4	38	0.56	0.48	0.71	31
1875292345_2_primary.phore	6	107	0.55	0.7	0.59	27
65432398_1_primary.phore	5	89	0.55	0.59	0.73	20
299784551_0_primary.phore	4	44	0.55	0.61	0.81	17
2325864244_1_primary.phore	4	93	0.54	0.76	0.75	20
9111456_2_PO.phore	7	122	0.53	0.57	0.69	27
247661025_0_PO.phore	4	81	0.52	0.8	0.57	17
823806273_2_primary.phore	5	134	0.51	0.6	0.69	23
193489750_1_primary.phore	5	103	0.51	0.6	0.7	26
23114756_0_primary.phore	4	84	0.51	0.67	0.79	16
2325792021_2_PO.phore	3	62	0.51	0.86	0.78	10
426406050_0_primary.phore	5	76	0.5	0.56	0.76	13
32933521_1_PO.phore	6	94	0.46	0.56	0.63	25

Supplementary Table 21 | The distributions of features and exclusion spheres in the pharmacophore models of the LigPhore test set, alongside the corresponding mapping scores, PoseBusters success rates, chemical diversity, and atom counts of generated molecules (part 9/9).

Pharmacophore	Feature count	Exclusion sphere count	Mapping score	PoseBusters success rate	Chemical diversity	Atom count
195793783_1_primary.phore	3	61	0.45	0.87	0.79	11
16940319_1_PO.phore	5	48	0.39	0.59	0.64	22
100985791_2_primary.phore	6	56	0.32	0.59	0.75	21
493826_2_primary.phore	3	61	0.26	0.87	0.85	9

Supplementary Table 22 | The comparison for reference molecules and molecules generated by PhoreGen and other baseline models.

Target	Reference	PhoreGen	AR	Pocket2Mol	DeepICL	TargetDiff	DiffSBDD	PMDM	DrugFlow	ShePhERD
TrmD										
JAK-2										
EBNA-1										
COP1										
Hsp90										
hCRBPII										
PP2C										
JMJD6										
UCK2										
HATs										

Supplementary Table 23 | Percentage of different ring sizes of the molecules generated by PhoreGen and baseline models on the 10 selected targets.

Ring size	Reference	AR[2]	Pocket2Mol[3]	DeepICL[4]	TargetDiff[5]	DiffSBDD[6]	PMDM[7]	DrugFlow[8]	ShEPhERD[9]	PGMG[10]	PhoreGen
3	0.0%	22.7%	0.0%	0.3%	0.0%	16.0%	0.2%	3.3%	15.3%	0.2%	3.3%
4	0.0%	0.3%	0.2%	0.1%	1.7%	2.4%	0.3%	0.6%	3.4%	0.2%	1.4%
5	26.7%	7.2%	15.9%	5.7%	26.9%	21.1%	23.3%	24.7%	24.8%	14.2%	25.0%
6	56.7%	47.4%	67.3%	76.7%	45.6%	33.8%	42.9%	52.2%	35.0%	68.3%	51.2%
7	0.0%	0.3%	0.8%	3.1%	8.1%	6.6%	13.9%	2.5%	3.8%	0.9%	4.2%
8	0.0%	0.0%	0.0%	0.4%	1.2%	1.3%	0.7%	0.4%	0.9%	0.1%	0.6%
9	0.0%	0.1%	0.0%	0.0%	0.4%	0.7%	0.1%	0.2%	0.0%	0.0%	0.1%
Fused[a]	13.3%	6.9%	13.4%	7.8%	9.4%	11.8%	6.9%	10.2%	11.3%	12.1%	9.3%
Polycyclic[b]	3.3%	15.1%	2.4%	6.0%	6.7%	6.3%	11.8%	5.8%	5.4%	4.0%	4.9%

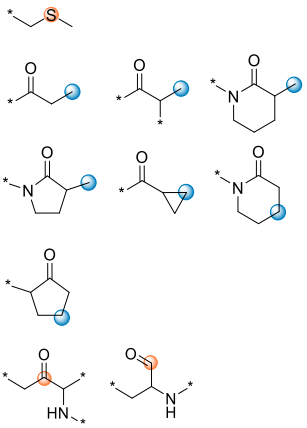
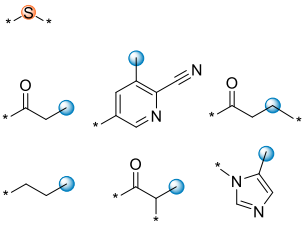
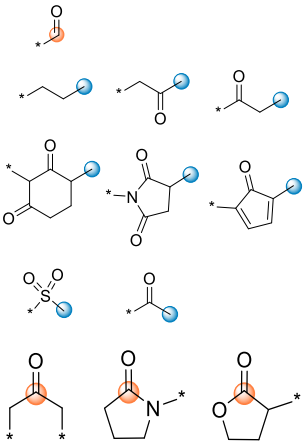
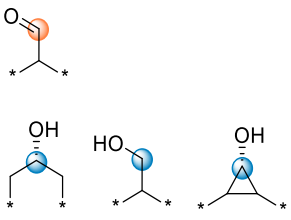
[a] Two rings fused. [b] More than two rings fused.

Supplementary Table 24 | The performance comparison of molecules generated by PhoreGen and baseline models on the CrossDocked2020 dataset (100 targets)[a].

Methods	QED (↑)	SA (↑)	Lipinsk (↑)	LogP	Diversity	Docking score (↓)	High affinity (↑)	IFP similarity (↑)	PoseBusters (↑)
Test set	0.48	0.73	4.27	0.89	-	-7.45	-	-	-
AR	0.51	0.64	4.75	0.39	0.68	-6.75	36.4%	0.57	0.61
Pocket2Mol	0.57	0.76	4.88	1.51	0.69	-7.15	47.7%	0.50	0.87
DeepICL	0.61	0.33	4.94	1.02	0.79	-7.04	45.9%	0.45	0.35
TargetDiff	0.48	0.58	4.51	1.36	0.72	-7.80	56.8%	0.58	0.57
DiffSBDD	0.47	0.58	4.56	1.22	0.72	-7.18	45.9%	0.54	0.50
PMDM	0.55	0.59	4.81	2.37	0.69	-7.24	54.9%	0.37	0.84
DrugFlow	0.52	0.72	4.60	1.13	0.72	-6.88	39.8%	0.63	0.80
ShEPhERD	0.53	0.67	4.83	0.78	0.71	-5.54	24.4%	0.62	0.92
PhoreGen	0.47	0.67	4.31	1.78	0.67	-7.49	49.7%	0.78	0.60

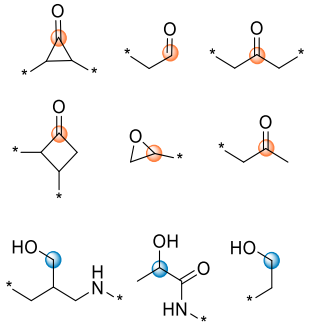
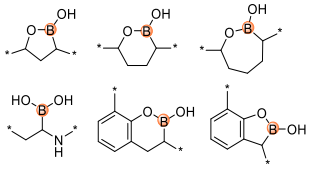
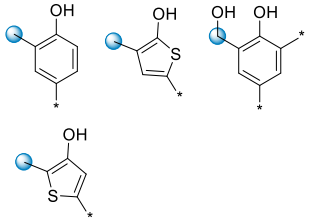
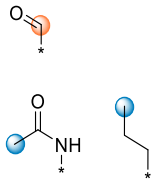
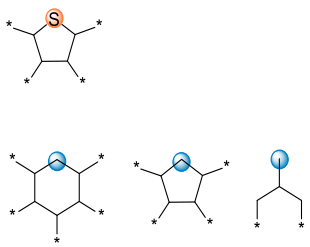
[a] The details regarding the 100 targets in CrossDocked2020 can be found in Ref. 11.

Supplementary Table 25 | PhoreGen-generated molecules with potential covalent motifs (part 1/2).

Target (UniProt ID)	PDB code	CV type	Covalent residue	Feature count	Representative generated covalent warhead[a]	Rate of reasonable CV	Average mapping score	Pre-/post-reaction counts
AKT1 (P31749)	6HHI	CV1	Cys_{296}	6		62%	0.708	1/61
Kgp (Q51817)	6I9A	CV1	Cys_{477}	7		99%	0.589	2/97
KRAS (P01116)	6P8Z	CV1	Cys_{12}	6		70%	0.754	2/68
GSTO1 (P78417)	6MHC	CV1	Cys_{32}	3		83%	0.549	1/82
PBP2 (P0AD65)	6G9F	CV2	Ser_{330}	3		92%	0.593	4/88

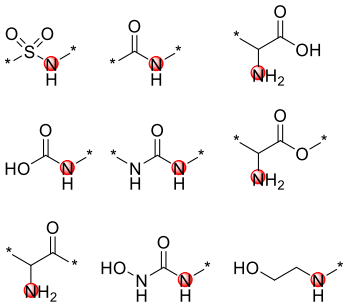
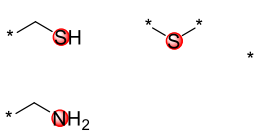
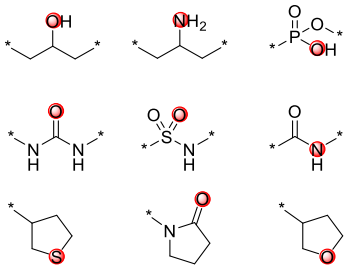
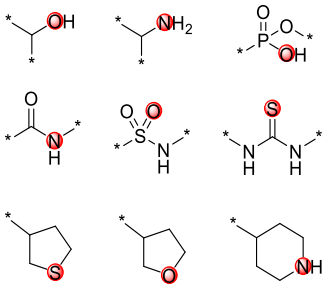
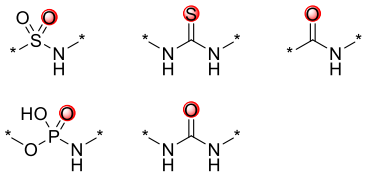
[a] The orange spheres represent the pre-reaction state of the covalent atom, while the blue spheres represent the post-reaction state.

Supplementary Table 26 | PhoreGen-generated molecules with potential covalent motifs (part 2/2).

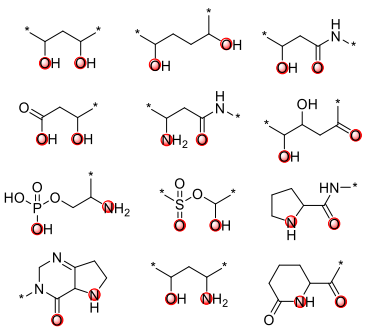
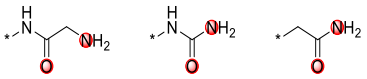
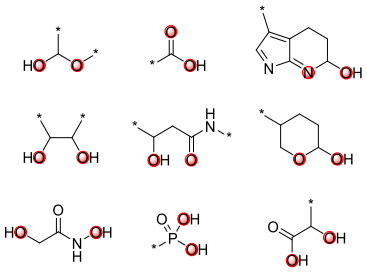
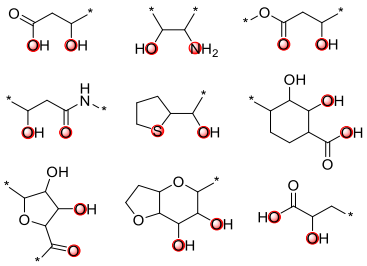
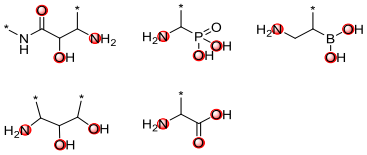
Target (UniProt ID)	PDB code	CV type	Covalent residue	Feature count	Representative generated covalent warhead[a]	Rate of reasonable CV	Average mapping score	Pre-/post-reaction counts
CTX-M-15 (G3G192)	6QW8	CV2	Ser_{70}	8		93%	0.640	11/82
OXA-10 (P14489)	6RTN	CV2	Ser_{67}	7		37%	0.754	37/0
ERN1 (Q9EQY0)	4PL5	CV3	Lys_{907}	7		62%	0.538	0/62
NDM-1 (E5KIY2)	6OVZ	CV3	Lys_{211}	2		63%	0.411	6/57
Ag111 (Q9STC1)	4AMX	CV4	Asp_{553}	5		97%	0.836	59/38

[a] The orange spheres represent the pre-reaction state of the covalent atom, while the blue spheres represent the post-reaction state.

Supplementary Table 27 | PhoreGen-generated molecules with potential metal-binding motifs (part 1/2).

Target (UniProt ID)	PDB code	Metal ion	MB count	Feature count	Representative MB motifs	Rate of reasonable MB	Average mapping score
HDAC10 (F1QCV2)	6UHU	Zn^{2+}	1	4		99%	0.409
JMJD6 (Q6NYC1)	6GDY	Fe^{3+}	1	4		95%	0.854
FTO (Q9C0B1)	4IE6	Zn^{2+}	1	5		100%	0.849
MMP3 (P08254)	1CAQ	Zn^{2+}	1	5		100%	0.849
NPR (P00800)	5N34	Zn^{2+}	1	5		100%	0.833

Supplementary Table 28 | PhoreGen-generated molecules with potential metal-binding motifs (part 2/2).

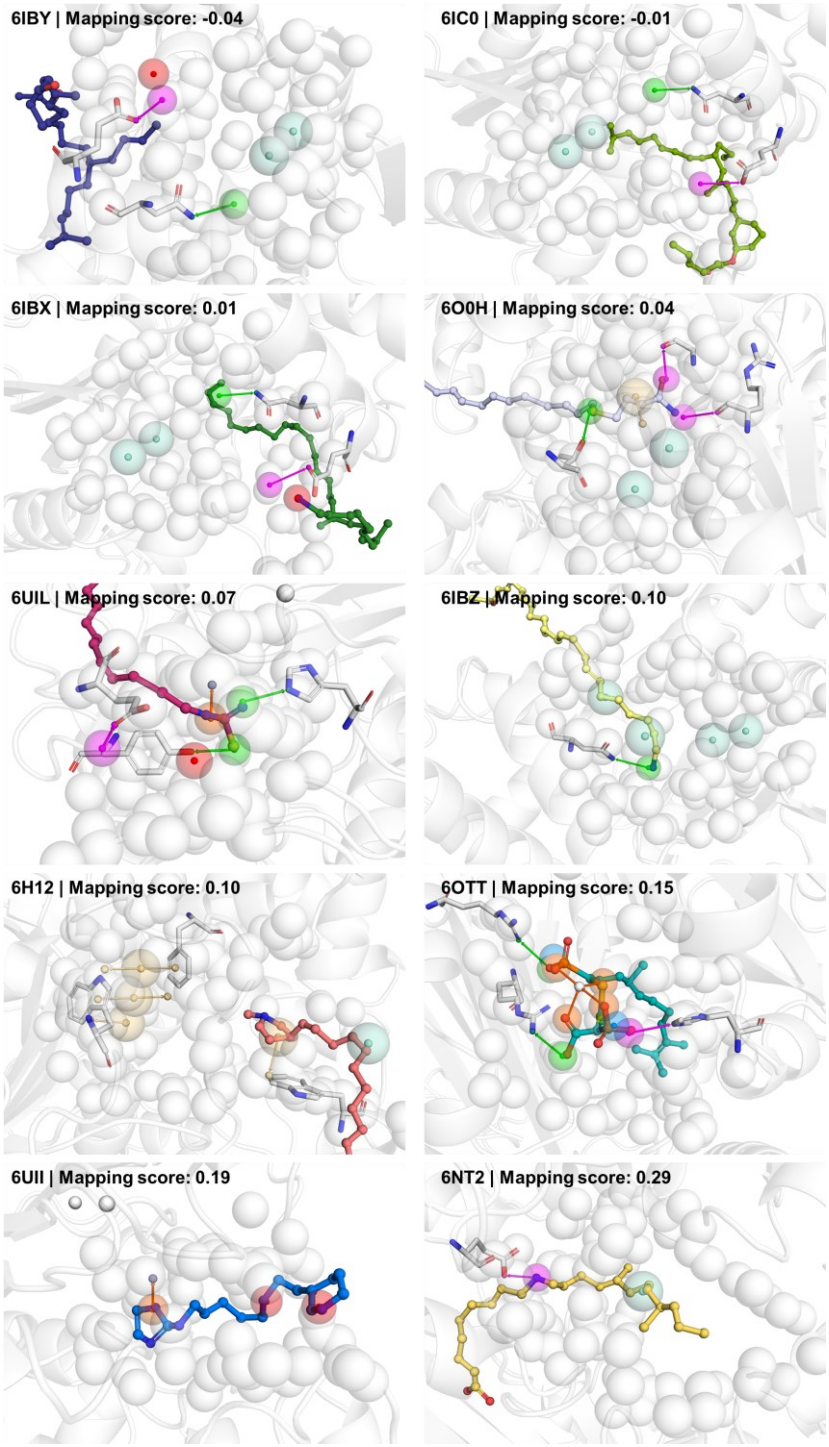
Target (UniProt ID)	PDB code	Metal ion	MB count	Feature count	Representative MB motifs	Rate of reasonable MB	Average mapping score
EGLN1 (Q9GZT9)	6ST3	Mn^{2+}	2	3		100%	0.719
MMP13 (P45452)	6HV2	Zn^{2+}	2	9		100%	0.690
HIF1AN (Q9NWT6)	2W0X	Fe^{3+}	2	6		100%	0.883
PA (C3W5S0)	6E4C	Mn^{2+}, Mn^{2+}	2	5		100%	0.880
MAP (P0AE18)	2GG8	Co^{2+}, Co^{2+}	4	6		100%	0.772

Supplementary Table 29 | Data collection and refinement statistics for OXA-48:**CB1** and *Dm*Naam:**CN2** complex structures.

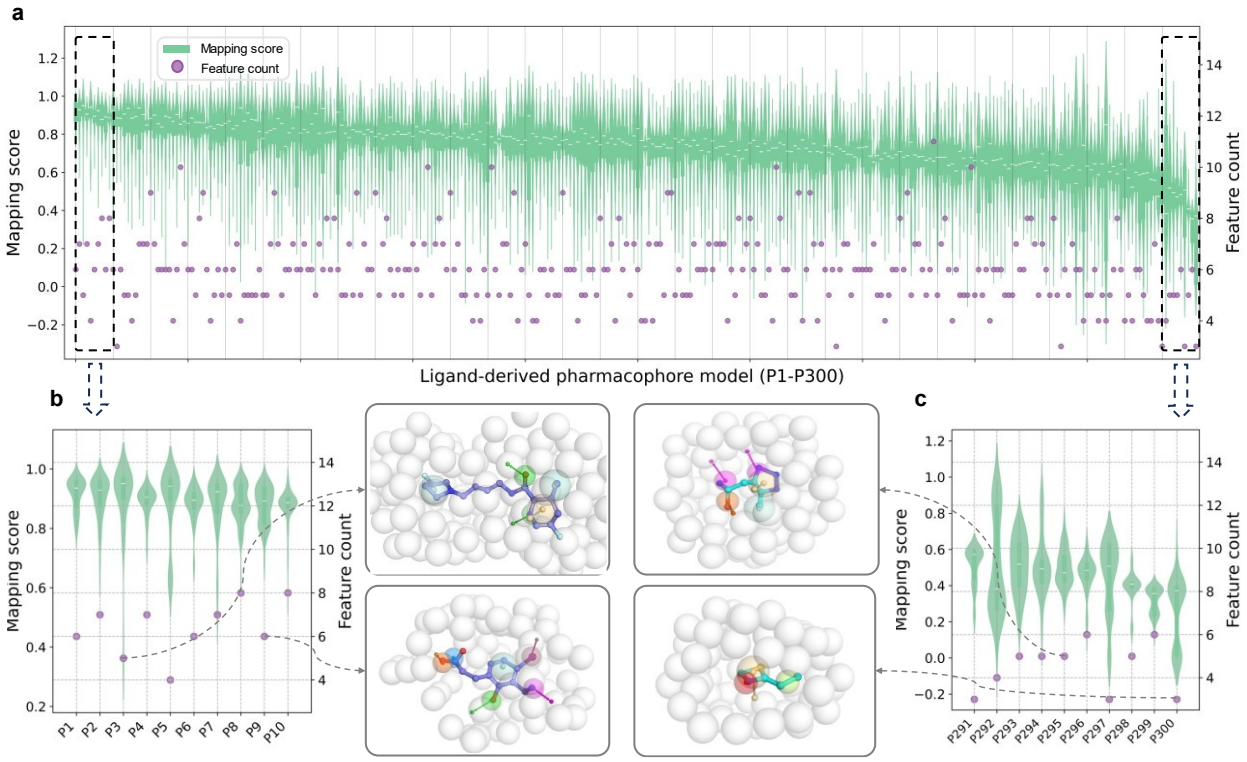
Structure	OXA-48:**CB1**	*Dm*Naam:**CN2**
PDB ID	9KSA	9U8M
Processing		
Radiation source	SSRF BL19U	SSRF BL19U
Space group	*P* 1211	*C* 1 2 1
Unit cell dimensions: a, b, c (Å)	104.60, 58.70, 105.86	74.863, 70.41, 76.524
Unit cell dimensions: α, β, γ (˚)	90.00, 104.79, 90.00	90, 94.459, 90
Number of molecules in ASU	4	1
Resolution range (outer shell) (Å)	38.401-2.178 (2.232-2.178)	41.51-1.758 (1.821~1.758)
Number of unique reflections	65,233	39,006
Completeness (%)	100.0	99.0
$I/\sigma(I)$ (outer shell)	4.16	10.74
R_{merge} (outer shell)	0.4276	0.1884
CC1/2	0.9956	0.9807
Wilson B factor (Å^2)	31.6	34.79
Refinement		
Overall B factor (Å^2)	45.30	44.66
Protein B factor (Å^2)	45.52	44.53
Ligand B factor (Å^2) (occupancy)	34.32 (1.0)	44.64 (1.0)
Water B Factor (Å^2)	41.78	47.26
RMSD from ideal bond length (Å)	0.015	0.011
RMSD from ideal angles (˚)	1.250	1.08
R_{work} (%)	20.51	20.17
R_{free} (%)	26.05	22.86

ASU: asymmetric unit; DmNaam: *Drosophila melanogaster* Naam; RMSD: root mean square deviation; SSRF: Shanghai Synchrotron Radiation Facility.

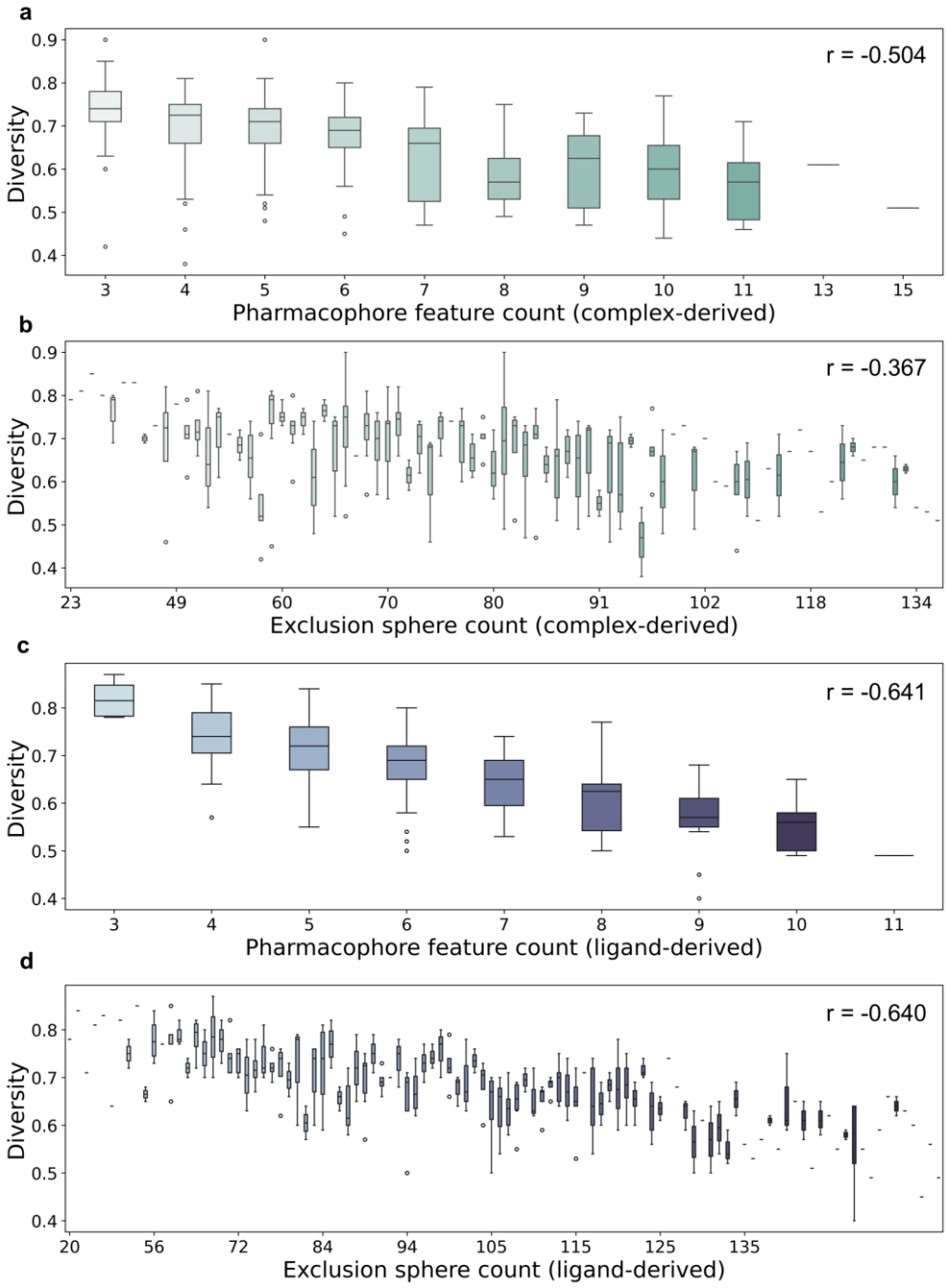
Section 3 Supplementary Figures



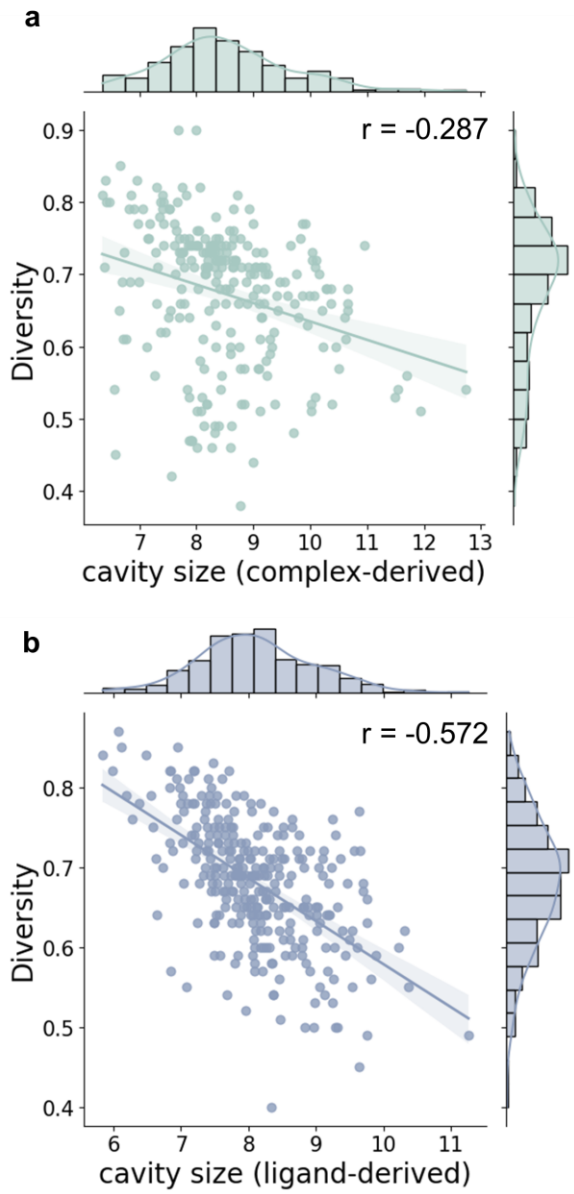
Supplementary Fig. 1 | View of the bottom-10 ranked generated molecules with their corresponding pharmacophore models. It reveals that for pharmacophore models with spatially dense features, homogenized features, and/or distant features, PhoreGen struggles to generate high-quality 3D molecules. Boron atoms, nitrogen atoms, oxygen atoms, phosphorus atoms, sulfur atoms, and chlorine atoms are pink, blue, red, darkorange, goldenrod, and limegreen respectively, while the other colors are carbon atoms.



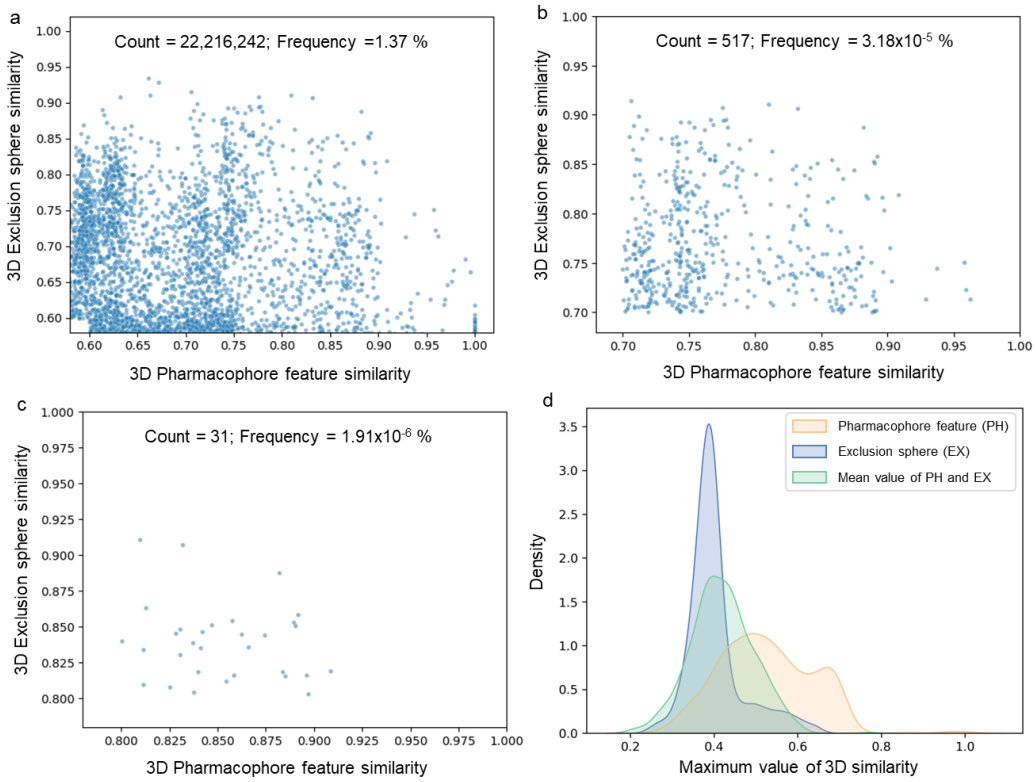
Supplementary Fig. 2 | Relationships between mapping scores of the generated molecules and feature counts of the ligand-derived pharmacophore models. a, Distributions of feature counts of the ligand-derived pharmacophore models and corresponding mapping scores of the generated molecules. **b,c** Detailed views are provided for the 10 ligand-derived pharmacophore models, showcasing the generated molecules with the highest (n = 100) and lowest (n = 100) mapping scores, respectively. Examples include high-scoring or low-scoring molecules generated for selected pharmacophore models. Carbon atoms are displayed as slateblue or cyan, and nitrogen atoms, oxygen atoms, and sulfur atoms are displayed as blue, red, and goldenrod, respectively. The box represents the interquartile range (IQR), the line inside the box is the median, the whiskers extend to ±1.5×IQR, and dots outside the whiskers are outliers.



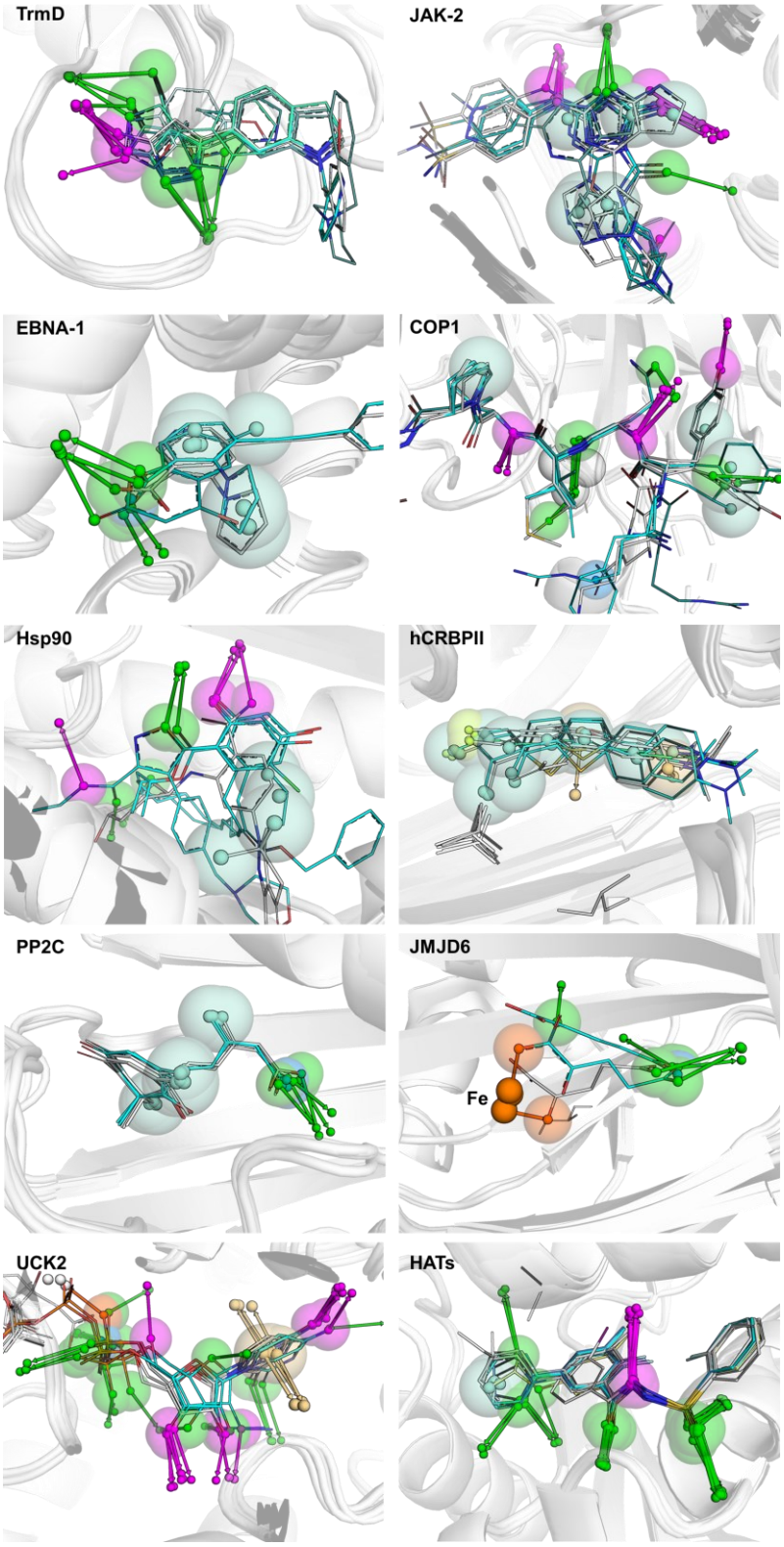
Supplementary Fig. 3 | Distributions of the pharmacophore feature and exclusion sphere counts versus the chemical diversity of generated molecules. **a**, **b** For the complex-derived pharmacophore models in the CpxPhore test set, the feature and exclusion sphere counts show a degree of correlation with the chemical diversity of the generated molecules. **c**, **d** For the ligand-derived pharmacophore models in the LigPhore test set, the feature and exclusion sphere counts also correlate with the chemical diversity in generation. Shadings indicate the feature or exclusion sphere counts, with darker shades representing higher counts. The box is the interquartile range (IQR), the line inside the box is the median, the whiskers represent data within 1.5×IQR, and the dots outside the whiskers are outliers. The r value refers to the Pearson correlation coefficient.



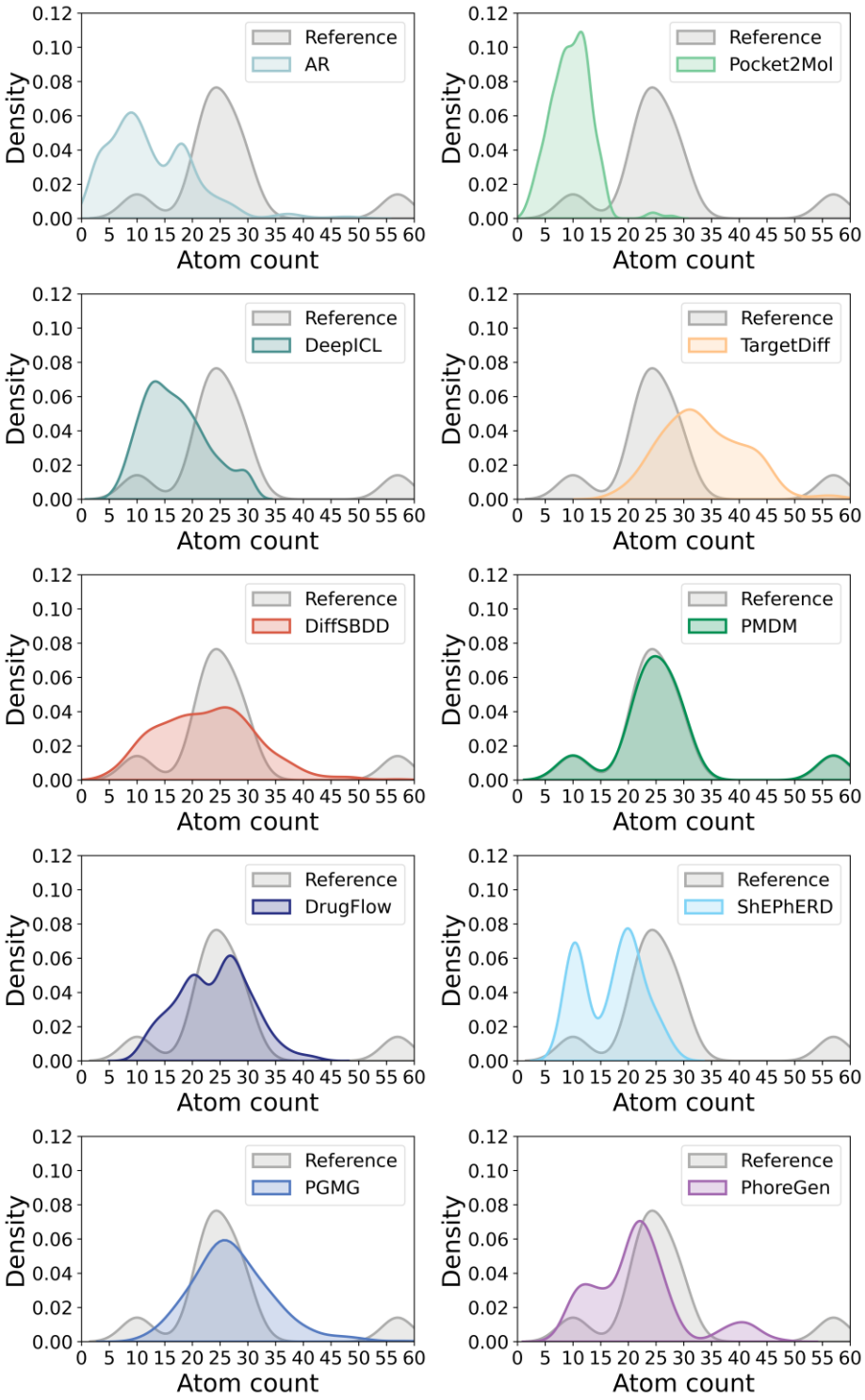
Supplementary Fig. 4 | Correlation between the cavity sizes of pharmacophore models and the chemical diversity of generated molecules. a, b, For both the complex-derived (**a**) and ligand-derived (**b**) pharmacophore models, the generated molecular diversity has a certain degree of correlation with the cavity sizes of pharmacophore models. The r value refers to the Pearson correlation coefficient.



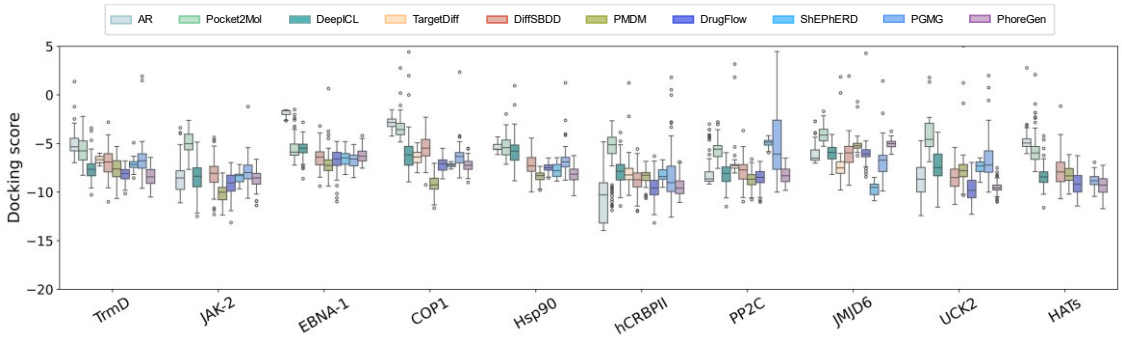
Supplementary Fig. 5 | The similarity analysis of 3D pharmacophore models between training and test set. This analysis was conducted on 665 pharmacophore models from the test set and 2,439,581 pharmacophore models from the training set, totaling 1,622,321,365 pairwise comparisons. The goal was to evaluate their 3D pharmacophore feature similarity and 3D exclusion sphere similarity. To manage the computational scale, we calculated pharmacophore fingerprint similarities (a 2D descriptor based on the type and number of pharmacophore features, where a similarity of 1 indicates identical feature types and counts) between the test and training set pharmacophore models, applying a threshold of 0.6, which reduced the number of comparisons to approximately 31% of the total. **a**, A total of 22,216,242 pairs (frequency: 1.37%) exhibit both pharmacophore feature similarity and exclusion sphere similarity above the threshold of 0.6. **b**, At a threshold of 0.7, only 517 pairs (frequency: 3.18×10^{-5} %) exceed both similarity metrics. **c**, At a threshold of 0.8, the number of pairs decreases to 31 (frequency: 1.91×10^{-6} %). d, The distribution of maximum 3D similarity values (mean of pharmacophore feature and exclusion sphere 3D similarities) for each test set pharmacophore compared to all training set pharmacophores shows that most values are around 0.4, with a maximum of 0.77. These results revealed the low 3D pharmacophore similarity between the training and test sets.



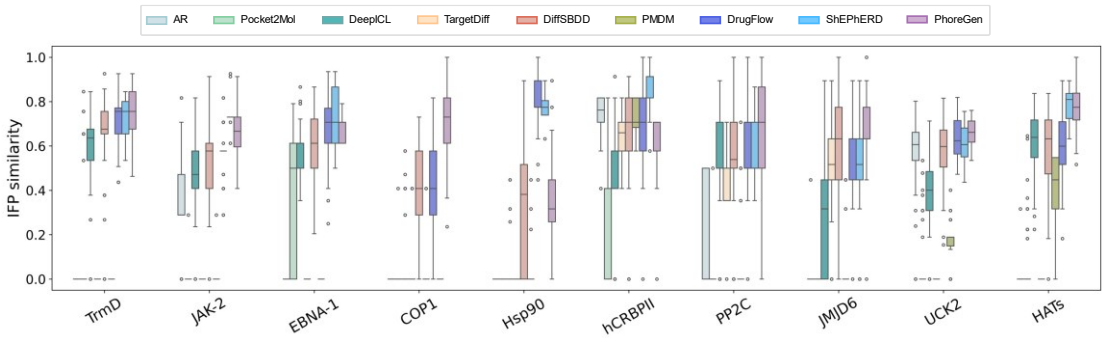
Supplementary Fig. 6 | The common interaction patterns and pharmacophore features of the 10 selected targets. For all these targets, the pharmacophore features are derived from complex crystal structures. Boron atoms, nitrogen atoms, oxygen atoms, phosphorus atoms, sulfur atoms, and chlorine atoms are pink, blue, red, darkorange, goldenrod, and limegreen respectively, while the other colors are carbon atoms.



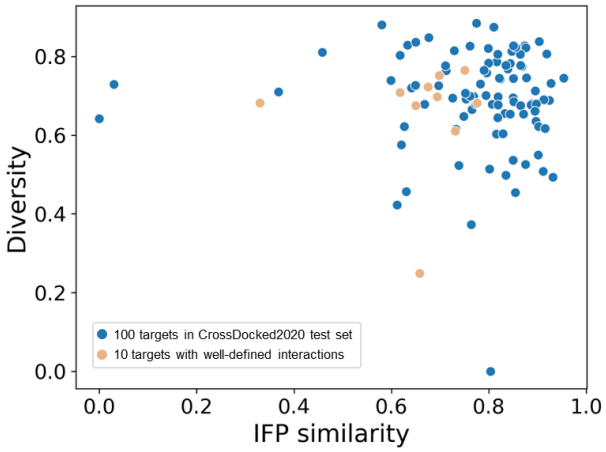
Supplementary Fig. 7 | The atom count distributions of the generated molecules versus target reference molecules. It reveals that for the 10 selected targets, the AR, Pocket2Mol, DeepICL, DiffSBDD, and ShEPhERD models tend to generate molecules with low molecular weights, while TargetDiff often generates molecules with relatively larger molecular weights than the reference molecules. PMDM, PGMG and PhoreGen effectively avoids generating molecules with undesirable characteristics, such as low molecular weights or sizes that deviate significantly from reference compounds.



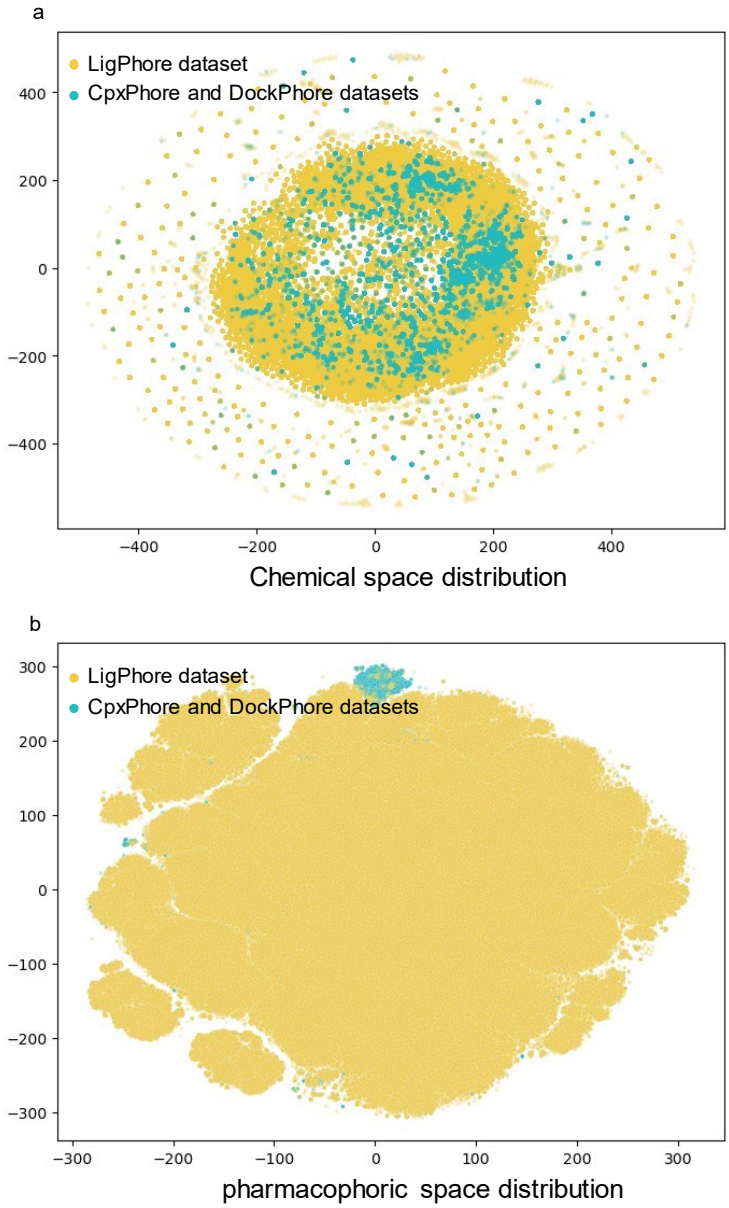
Supplementary Fig. 8 | Distributions of the docking scores for the molecules generated by the five generative models across the 10 selected targets. It reveals that PhoreGen shows a notable advantage over the nine baseline methods in generating molecules with potentially high binding affinity. The box is the interquartile range (IQR), the line inside the box is the median, the whiskers represent data within 1.5×IQR, and the dots outside the whiskers are outliers.



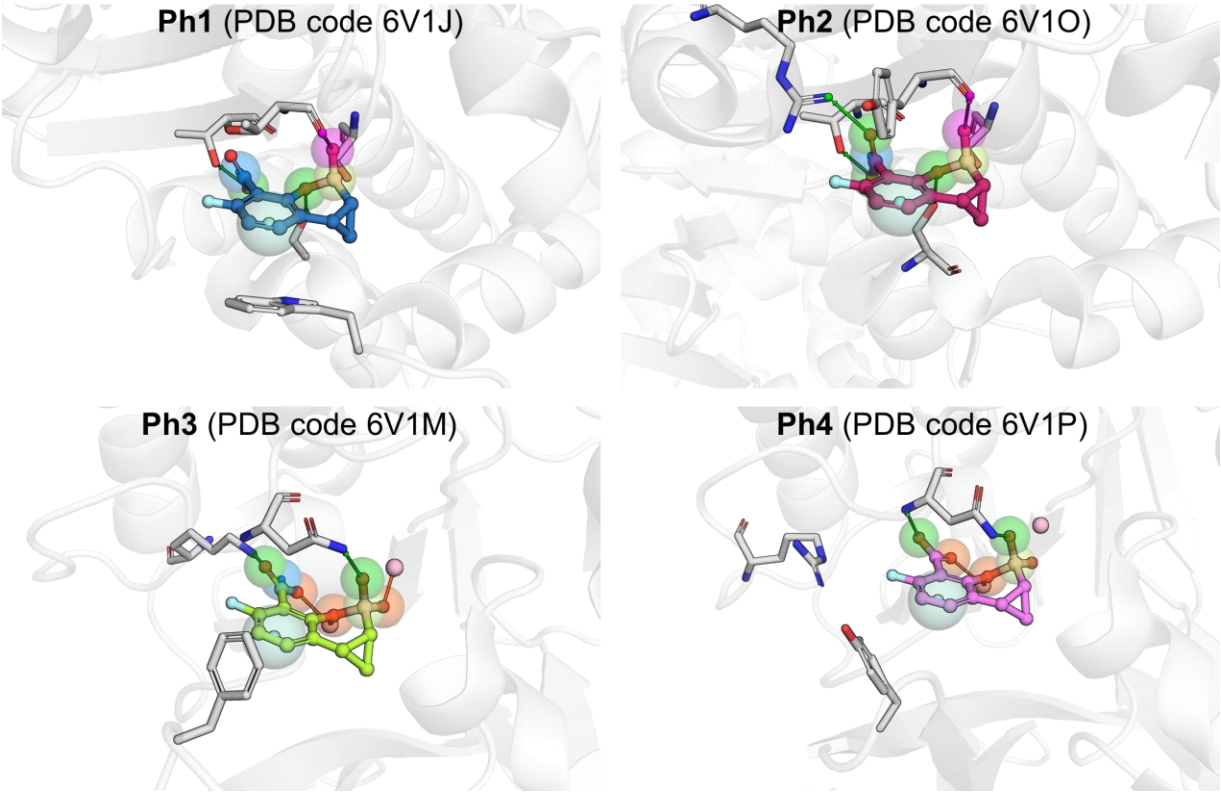
Supplementary Fig. 9 | Distributions of the IFP similarity scores for the molecules generated by the five generative models across the 10 selected targets. It reveals that PhoreGen-generated molecules closely align with known interaction patterns observed in these targets, suggesting a higher likelihood of being active molecules. The box is the interquartile range (IQR), the line inside the box is the median, the whiskers represent data within 1.5×IQR, and the dots outside the whiskers are outliers.



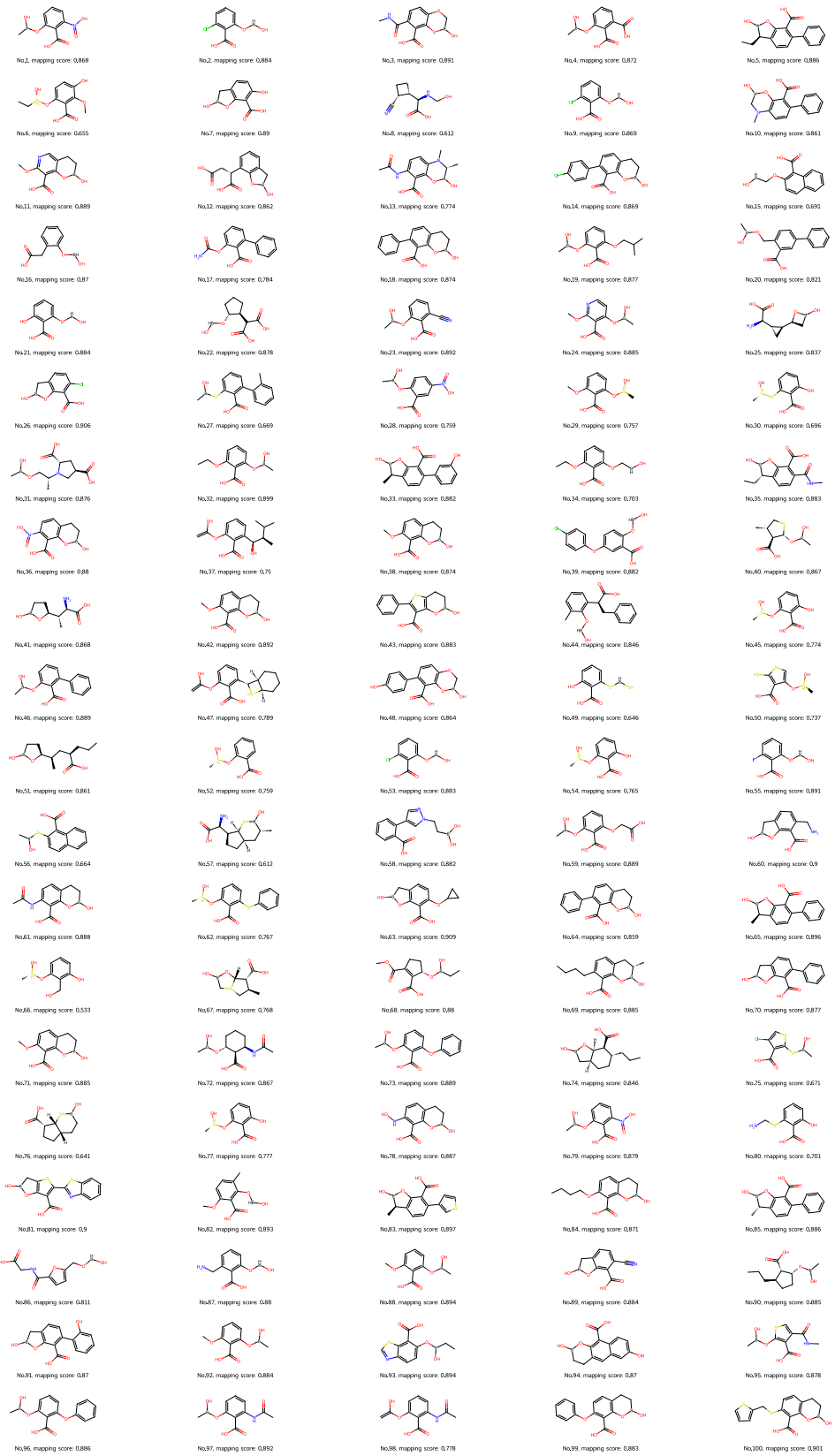
Supplementary Fig. 10 | Correlation between IFP similarity and chemical diversity of generated molecules. For both the 100 targets from the CrossDocked2020 test set and the 10 targets with well-defined interactions, the chemical diversity of the generated molecules remains substantial even with high IFP similarity.



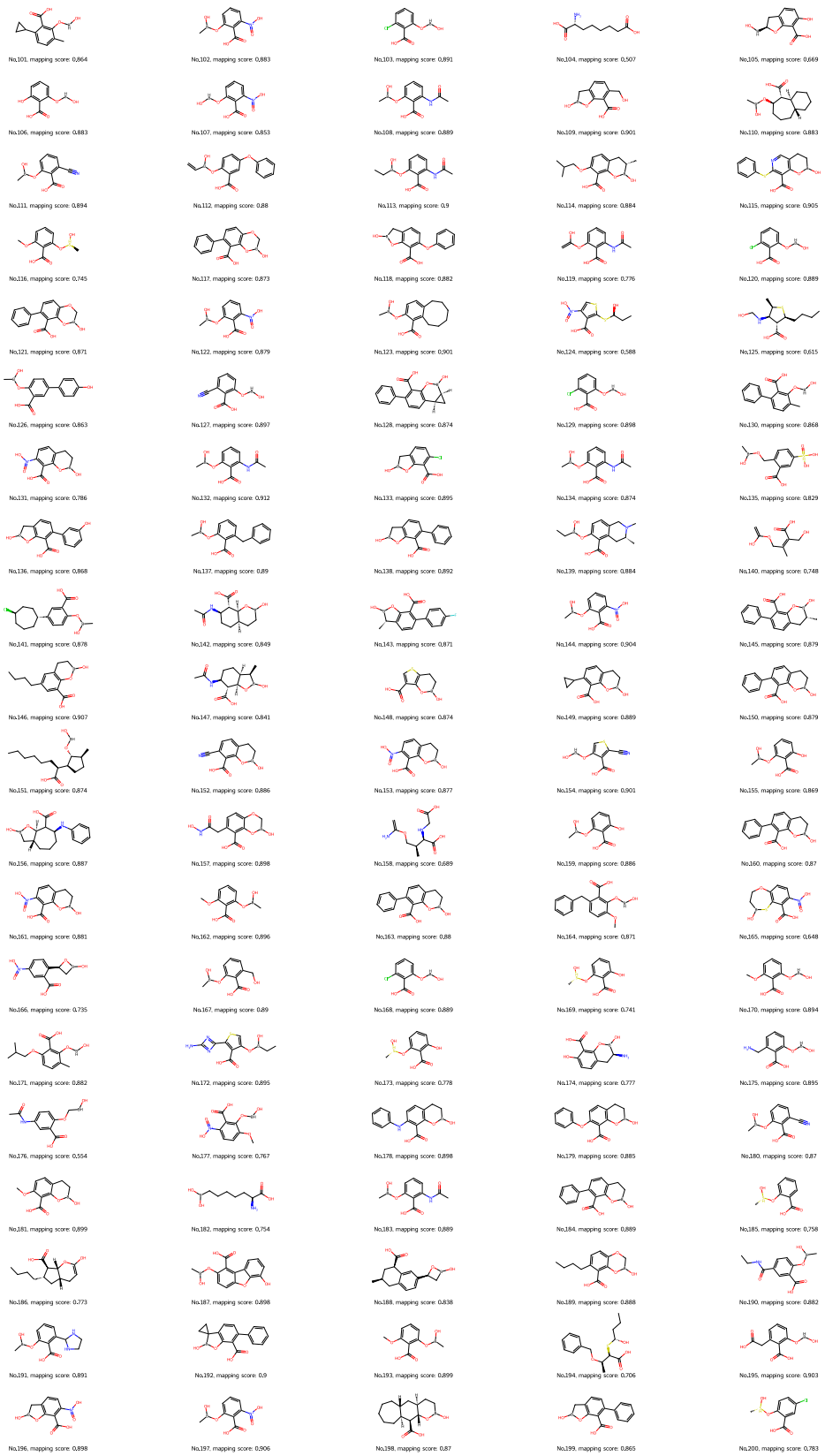
Supplementary Fig. 11 | The t-SNE analysis of chemical and pharmacophoric space distributions for ligand- and complex-derived datasets. The chemical space represented by ligands' ECFP4 (1024-bit) fingerprints (**a**) and the pharmacophoric space represented by pharmacophore feature counts (**b**) reveal that ligand-derived pharmacophores (*i.e.*, LigPhore dataset) and complex-derived pharmacophores (*i.e.*, CpxPhore and DockPhore datasets) exhibit both similarities and differences within the feature space. In comparison, ligand-derived pharmacophores show a broader distribution of feature diversity.. The ECFP4 fingerprints were processed by PCA (random_state=2024, n_components=50) for dimensionality reduction prior to the t-SNE analysis, which was performed with the following hyperparameters: n_components=2, perplexity=30, n_iter=5000, random_state=2024.



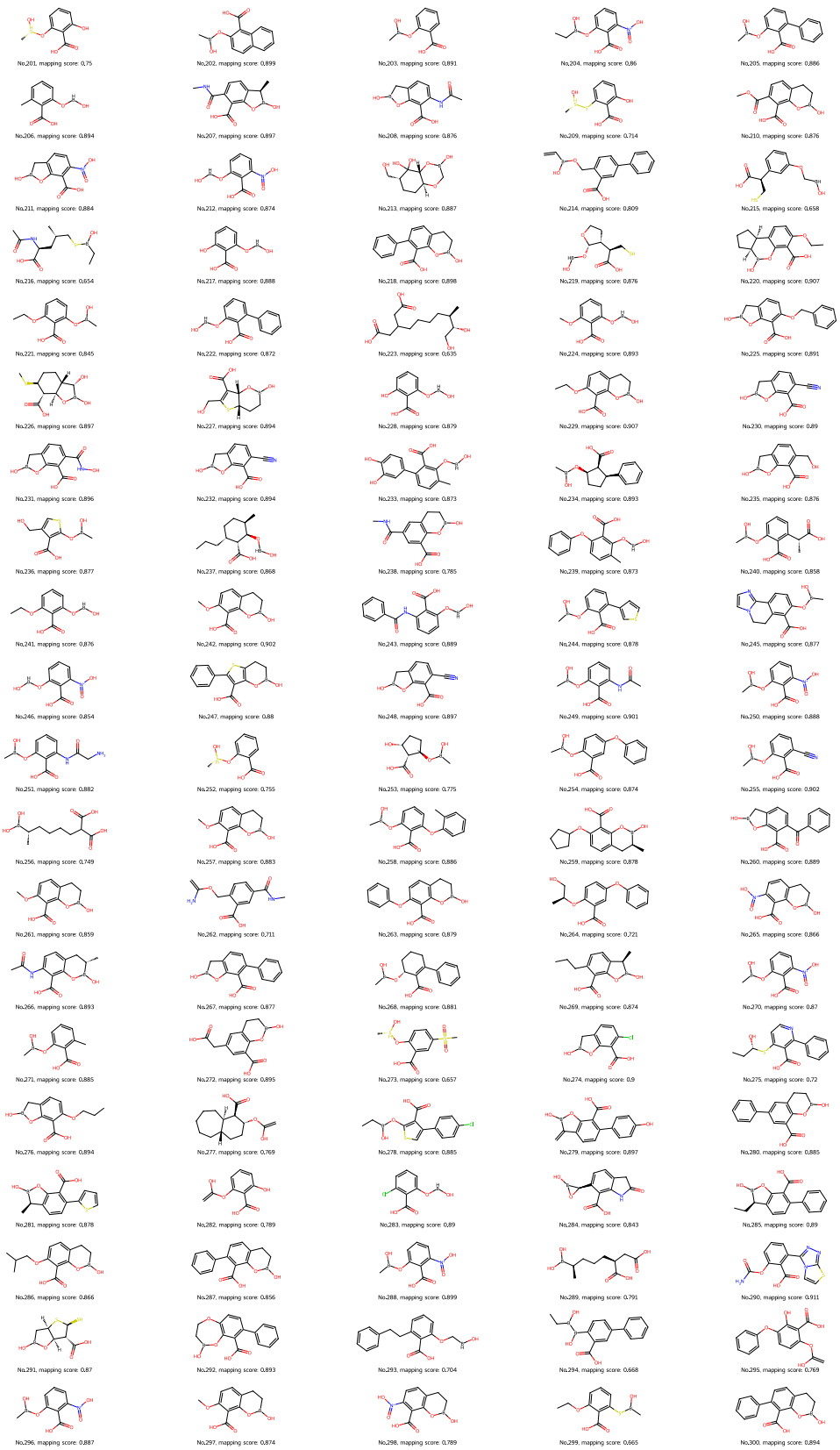
Supplementary Fig. 12 | The pharmacophore models derived from MBL/SBL complex crystal structures. The KPC-2:**QPX-7728** (PDB code 6V1J)[12], OXA-48:**QPX-7728** (PDB code 6V1O)[12], NDM-1:**QPX-7728** (PDB code 6V1M)[12], and VIM-2:**QPX-7728** (PDB code 6V1P)[12] complex structures were used for establishing the pharmacophore models **Ph1**-**Ph4**. Boron atoms, nitrogen atoms, oxygen atoms, and sulfur atoms are pink, blue, red, and goldenrod, respectively, while the other colors are carbon atoms.



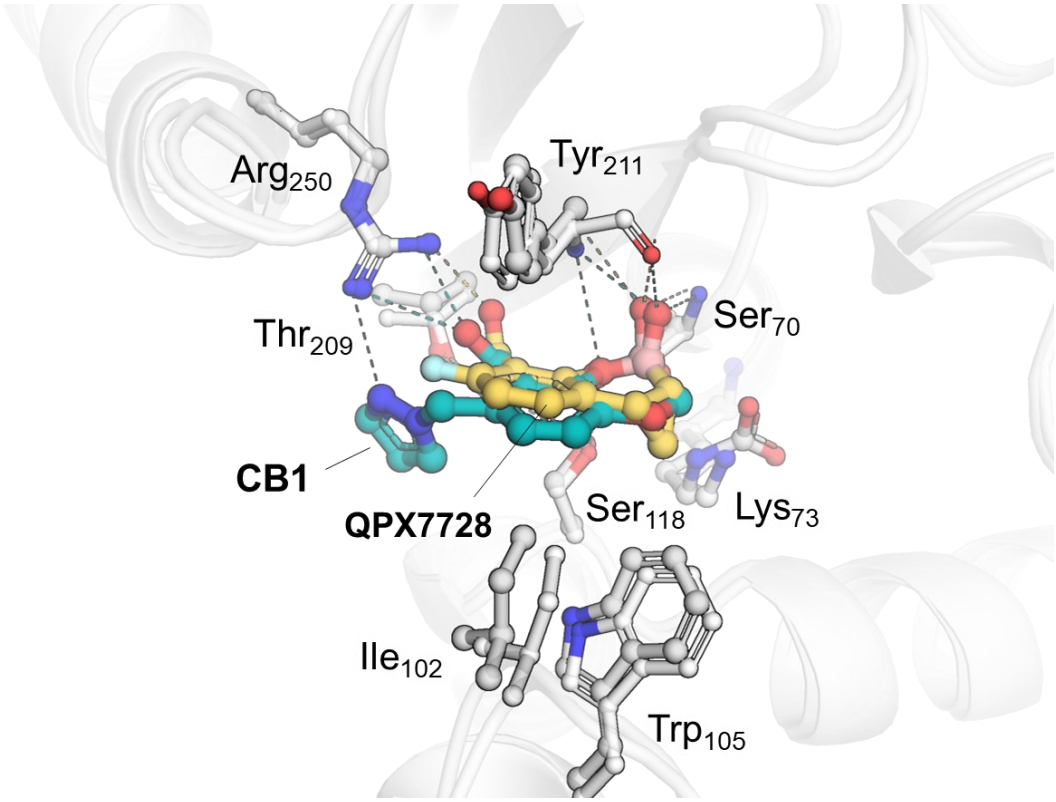
Supplementary Fig. 13 | The PhoreGen-generated molecules for MBLs/SBLs. The chemical structures of 300 generated molecules (part 1/3) using the pharmacophore model **Ph1** in Fig. 11 for molecular generation show that PhoreGen can generate structurally diverse boronates and other chemotypes for MBLs/SBLs.



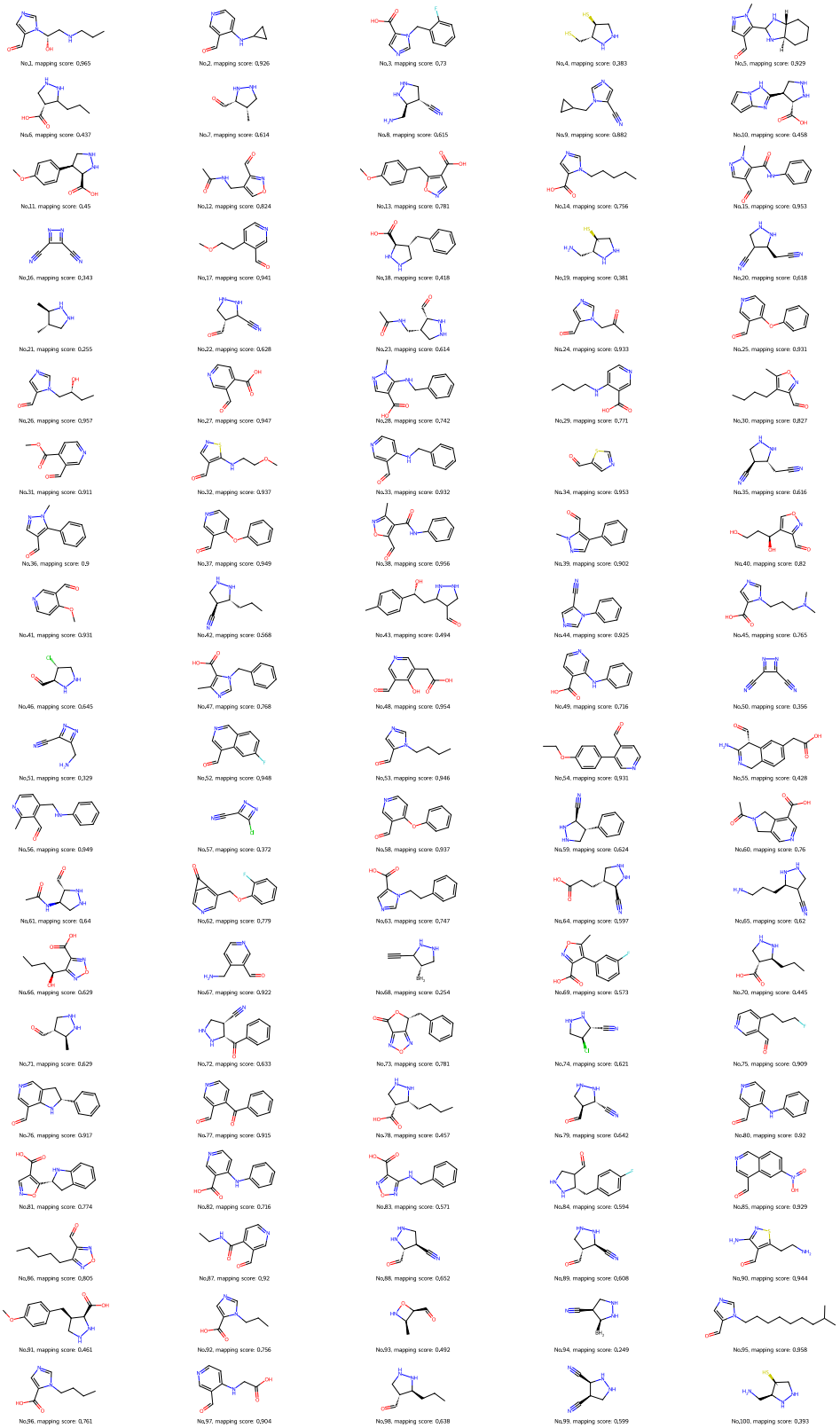
Supplementary Fig. 14 | The PhoreGen-generated molecules for MBLs/SBLs. The chemical structures of 300 generated molecules (part 2/3) using the pharmacophore model **Ph1** in Fig. 11 for molecular generation show that PhoreGen can generate structurally diverse boronates and other chemotypes for MBLs/SBLs.



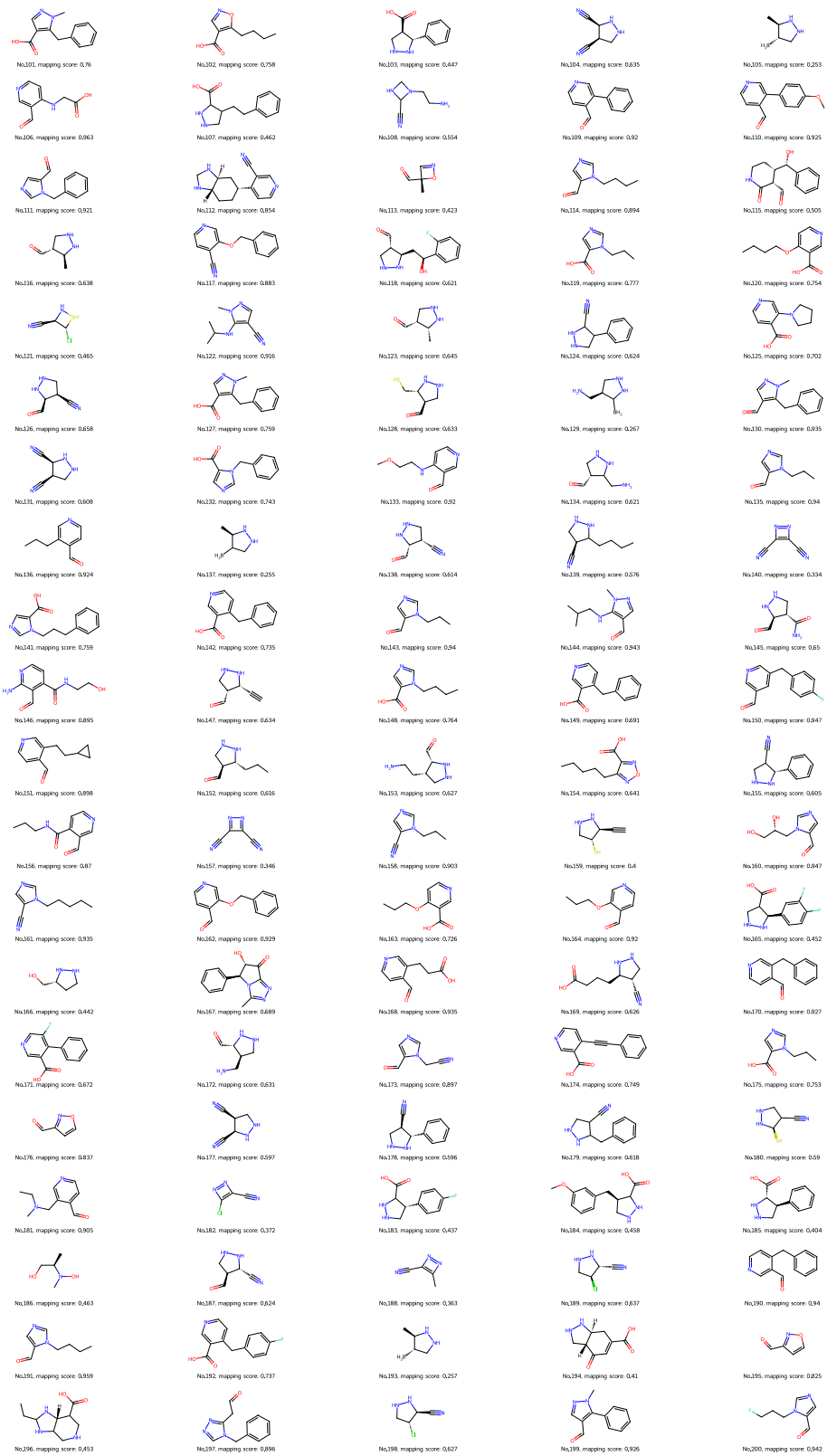
Supplementary Fig. 15 | The PhoreGen-generated molecules for MBLs/SBLs. The chemical structures of 300 generated molecules (part 3/3) using the pharmacophore model **Ph1** in Supplementary Fig. 12 for molecular generation show that PhoreGen can generate structurally diverse boronates and other chemotypes for MBLs/SBLs.



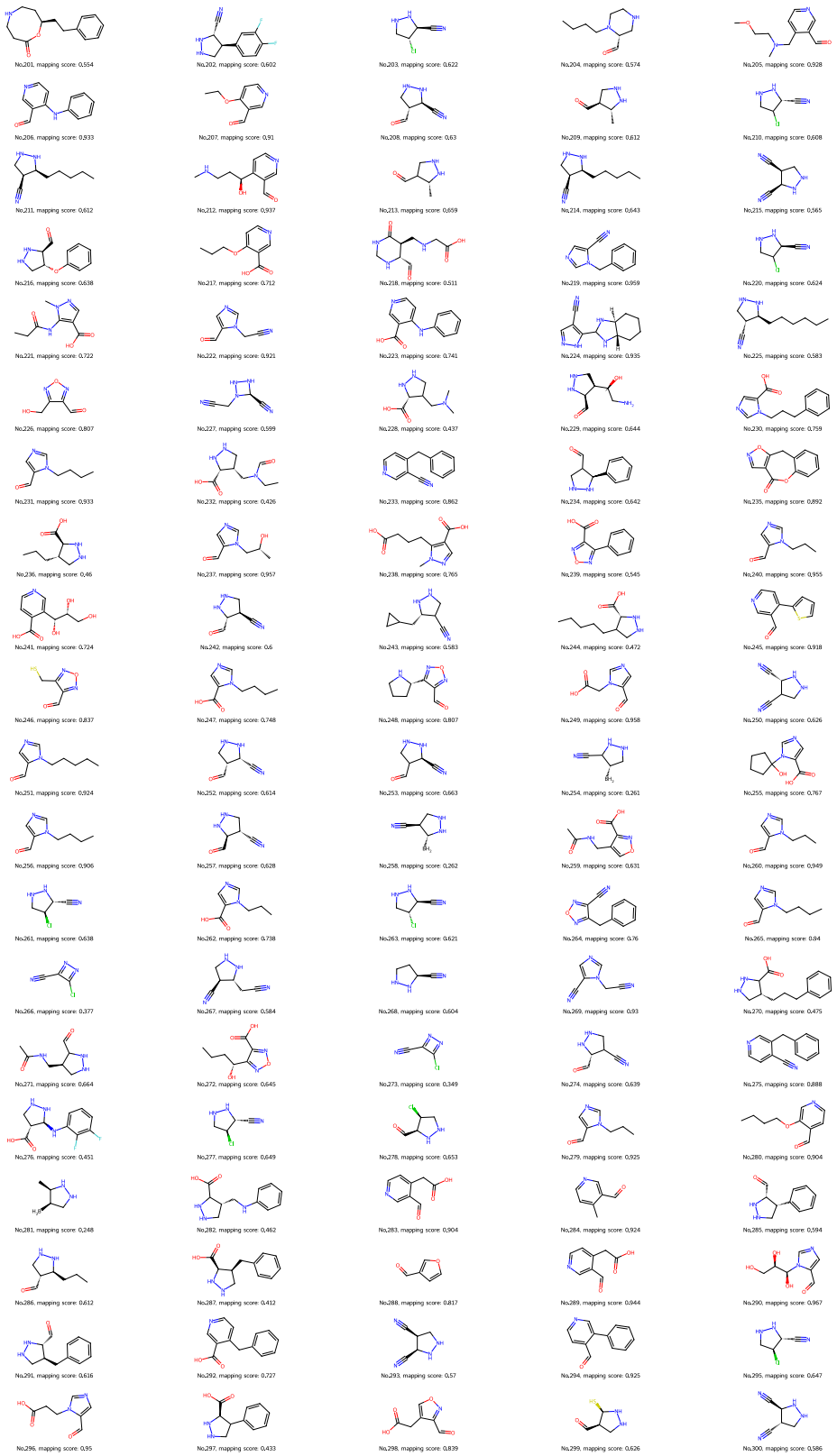
Supplementary Fig. 16 | CB1 and QPX7728 have a similar binding mode with OXA-48. Superimposition of OXA-48:**CB1** (PDB code 9KSA) and OXA-48:**QPX7728** (PDB code 6V1O)[12] reveals their similar binding modes with catalytically important residues, highlighting the effectiveness of PhoreGen in feature-customized drug discovery. Boron atoms, nitrogen atoms, oxygen atoms, and sulfur atoms are pink, blue, red, and goldenrod, respectively, while the other colors are carbon atoms.



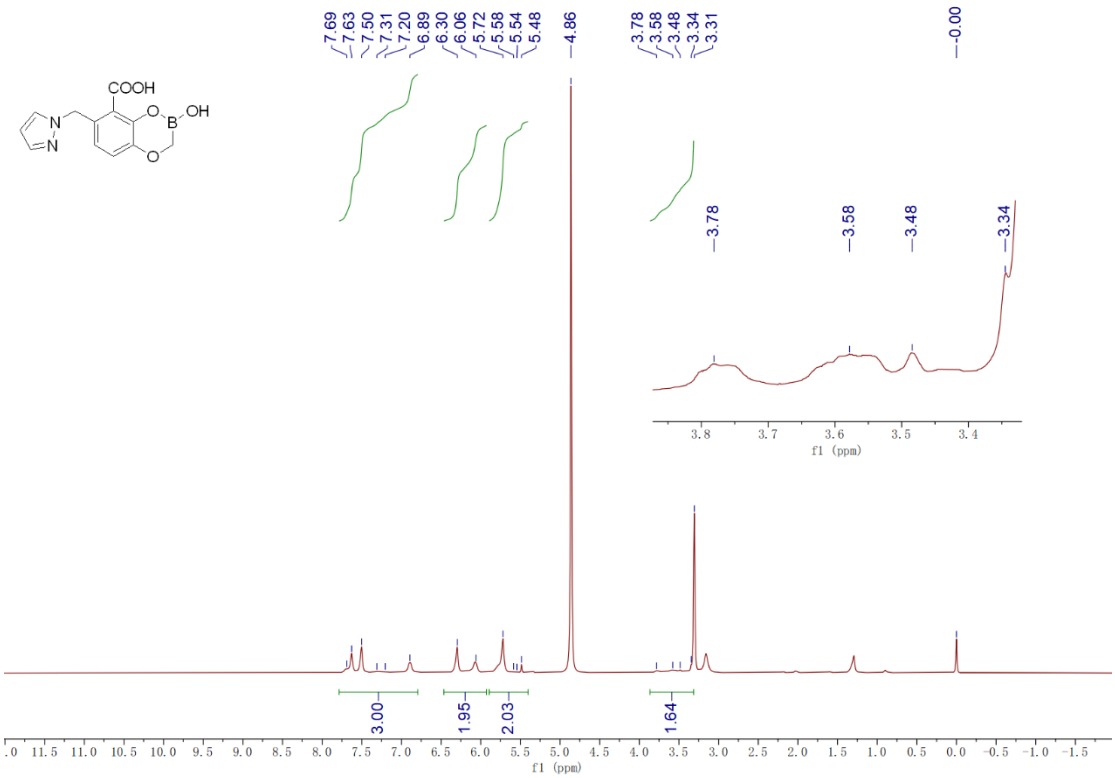
Supplementary Fig. 17 | The PhoreGen-generated molecules for Naam. The chemical structures of 300 generated molecules (part 1/3) using the pharmacophore model derived from Naam-catalytic mechanism for molecular generation show that PhoreGen can generate structurally diverse chemotypes with covalent groups for Naam.



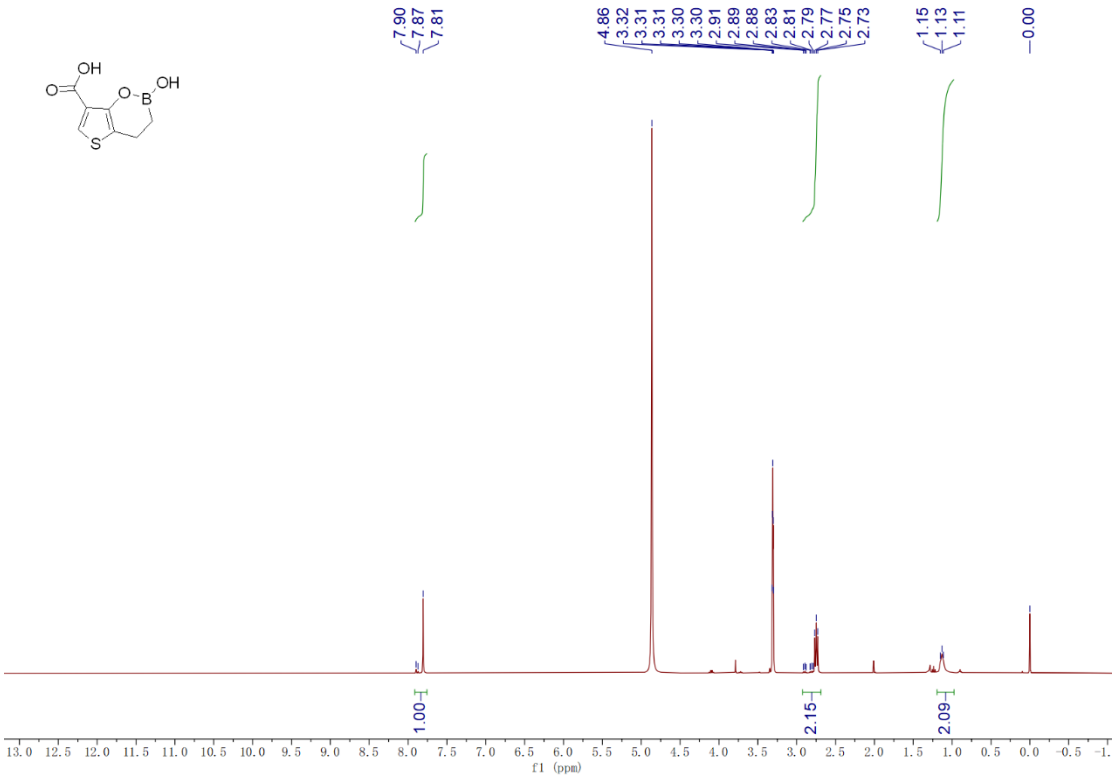
Supplementary Fig. 18 | The PhoreGen-generated molecules for Naam. The chemical structures of 300 generated molecules (part 2/3) using the pharmacophore model derived from Naam-catalytic mechanism for molecular generation show that PhoreGen can generate structurally diverse chemotypes with covalent groups for Naam.



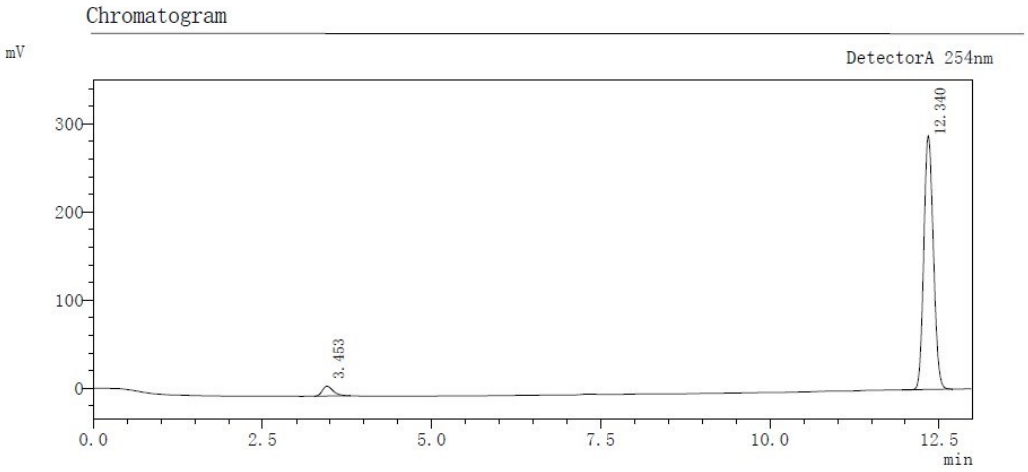
Supplementary Fig. 19 | The PhoreGen-generated molecules for Naam. The chemical structures of 300 generated molecules (part 3/3) using the pharmacophore model derived from Naam-catalytic mechanism for molecular generation show that PhoreGen can generate structurally diverse chemotypes with covalent groups for Naam.



Supplementary Fig. 20 | ^{1}H-NMR spectrum of compound CB1. ^{1}H NMR (400 MHz, CD_3OD) *δ* 7.79 – 6.79 (m, 3H), 6.18 (m, 2H), 5.60 (m, 2H), 3.86 – 3.31 (m, 2H).

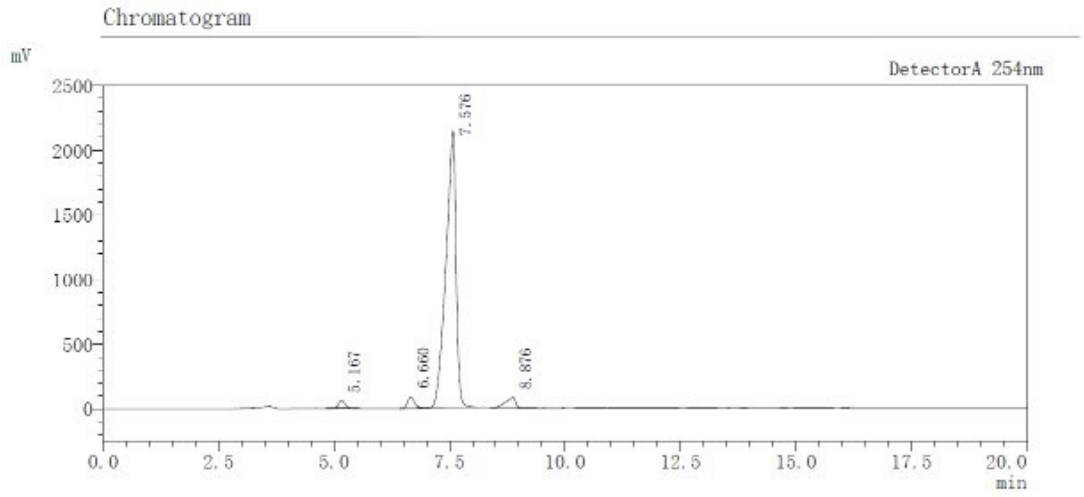


Supplementary Fig. 21 | ^{1}H-NMR spectrum of compound CB2. ^{1}H NMR (400 MHz, CD_3OD) δ 7.81 (m, 1H), 2.92 – 2.69 (m, 2H), 1.19 – 0.98 (m, 2H).



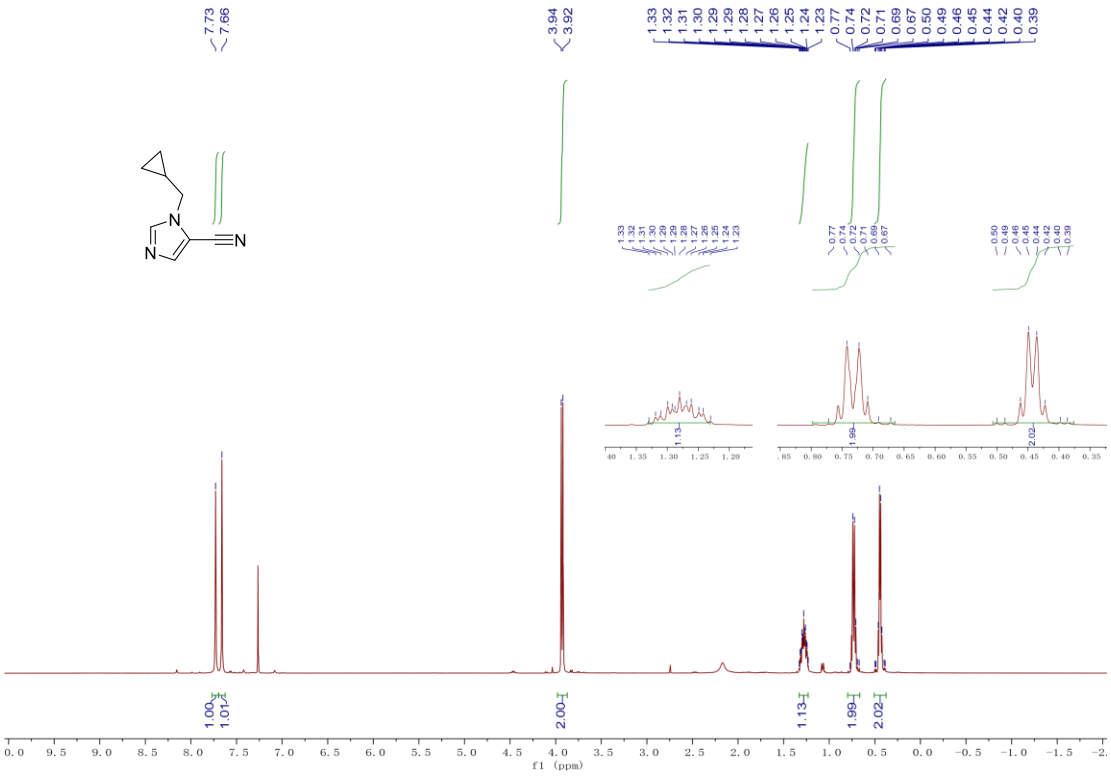
Peak information (PDA Ch1 254nm)			
Peak	Retention time	Peak area	Peak area %
1	3.453	121960	4.131
2	12.340	2830306	95.869
In total		2952266	100.000

Supplementary Fig. 22 | HPLC chromatogram of compound CB1. The HPLC purity is 95.869%.

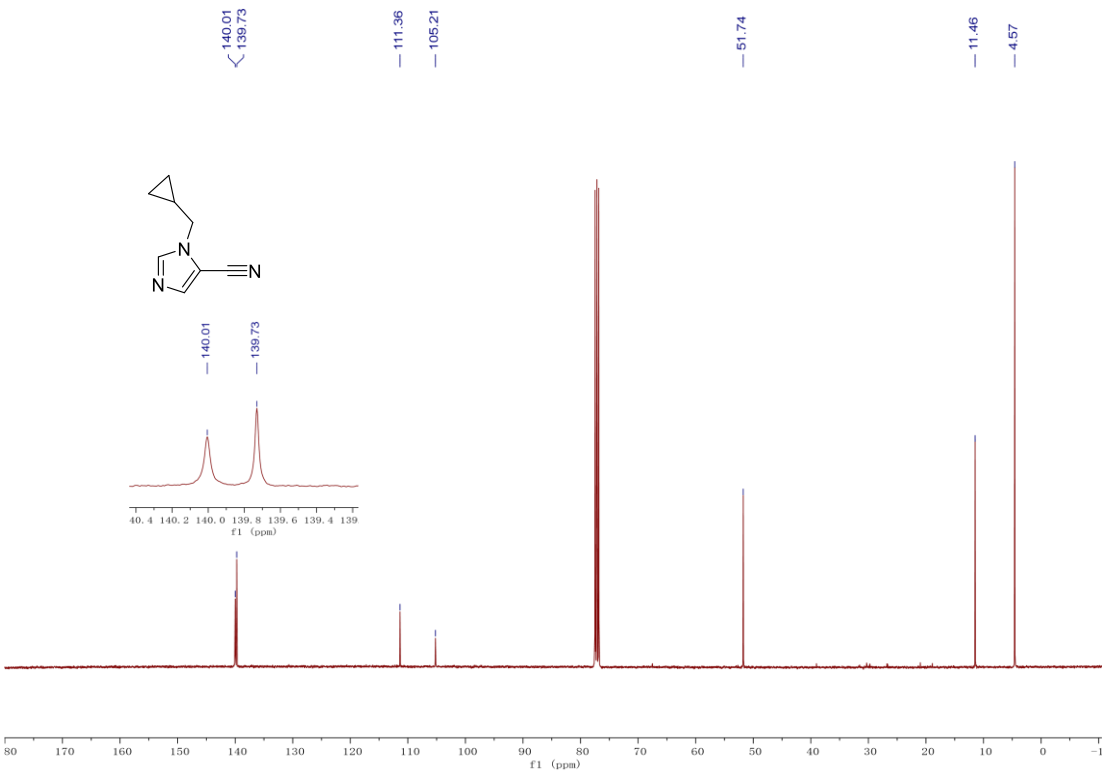


Peak information (PDA Ch1 254nm)			
Peak	Retention time	Peak area	Peak area %
1	5.167	663027	1.862
2	6.660	971349	2.727
3	7.576	32717883	91.862
4	8.876	1264024	3.549
In total		35616283	100.000

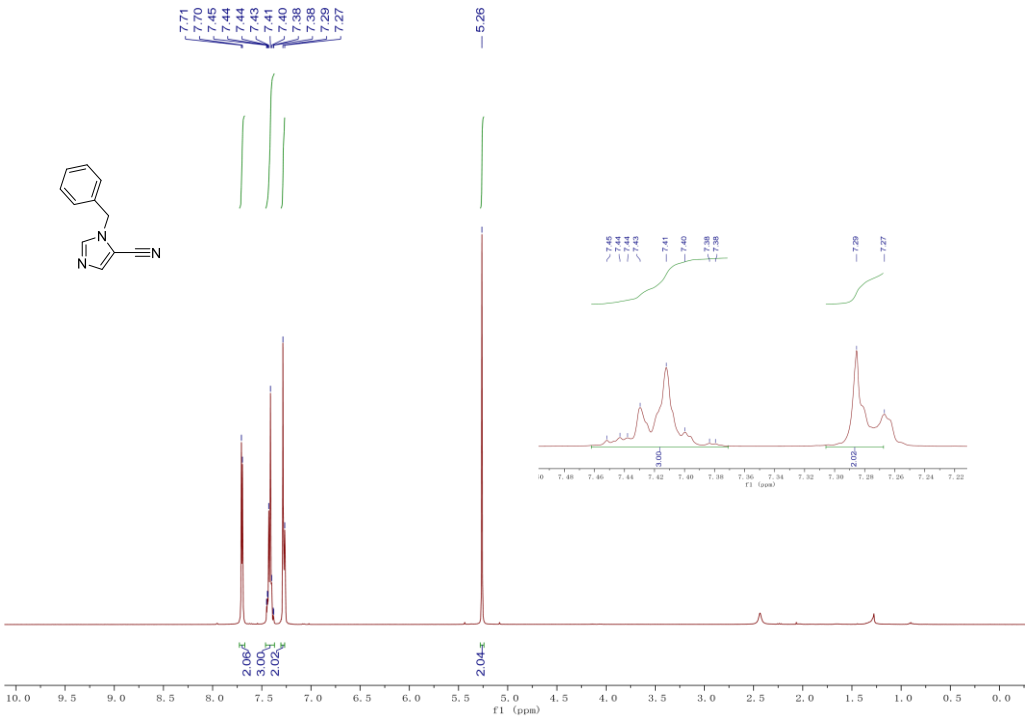
Supplementary Fig. 23 | HPLC chromatogram of compound CB2. The HPLC purity is 91.862%.



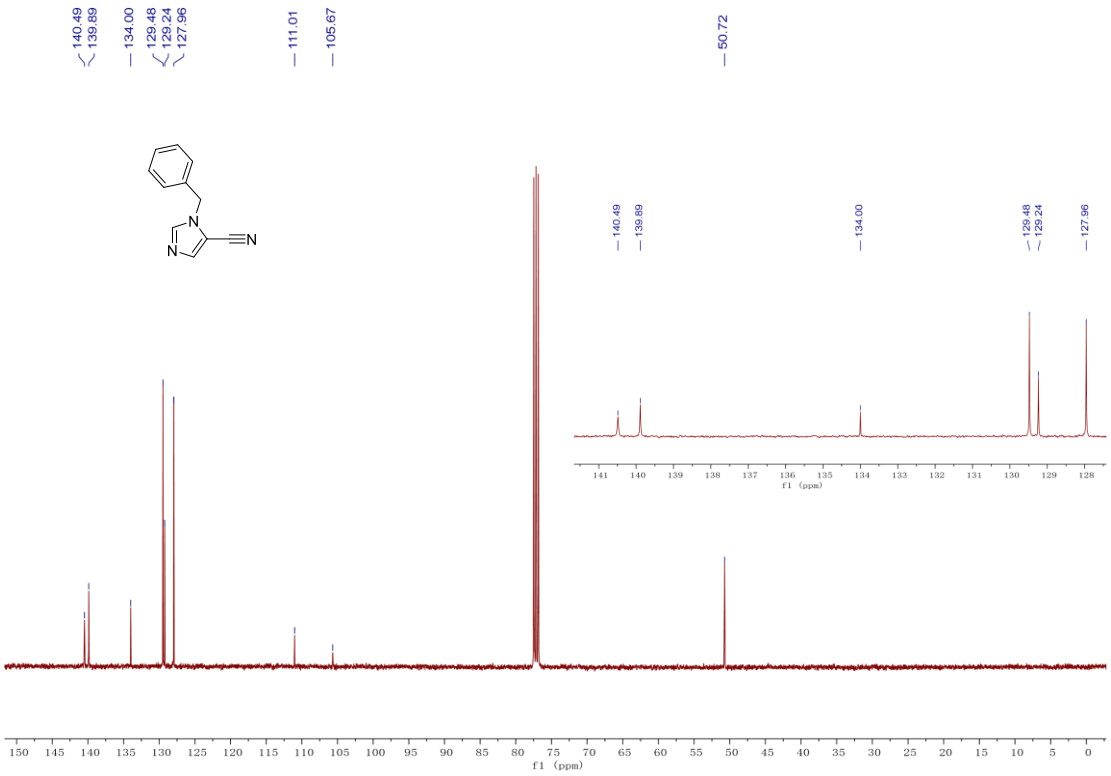
Supplementary Fig. 24 | ^{1}H-NMR spectrum of compound CN1. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.73 (s, 1H), 7.66 (s, 1H), 3.93 (d, J = 7.2 Hz, 2H), 1.22-1.34 (m, 1H), 0.80 – 0.66 (m, 2H), 0.44 (m, 2H).



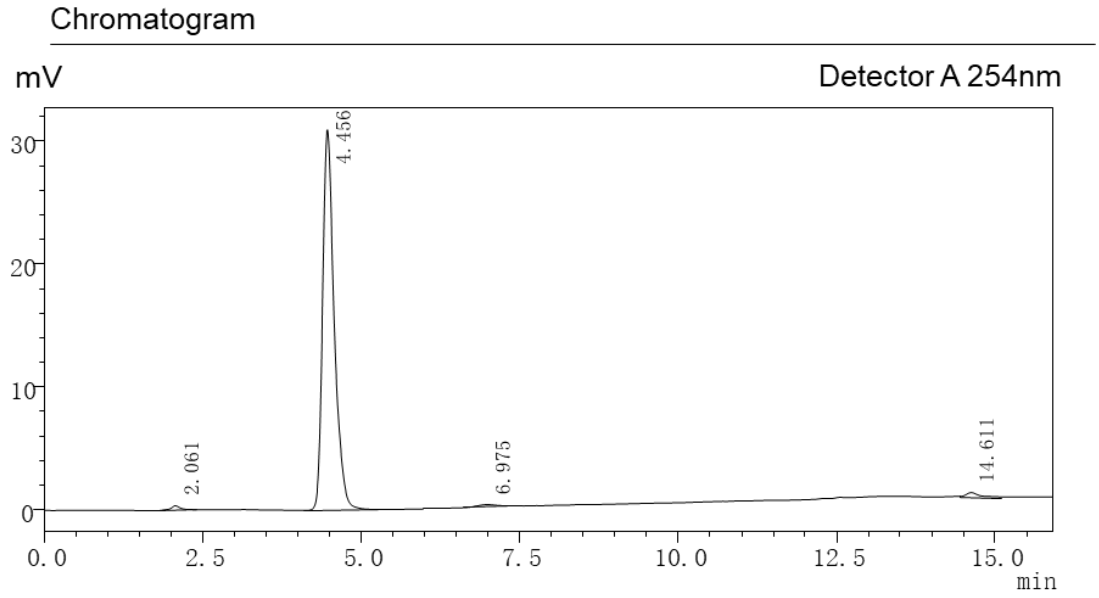
Supplementary Fig. 25 | ^{13}C -NMR spectrum of compound CN1. ^{13}C NMR (101 MHz, $CDCl_3$) δ 140.01, 139.73, 111.36, 105.21, 51.74, 11.46, 4.57.



Supplementary Fig. 26 | ^{1}H-NMR spectrum of compound CN2. ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 7.70 (d, J = 4.2 Hz, 2H), 7.46 – 7.37 (m, 3H), 7.24-7.32 (m, 2H), 5.26 (s, 2H).

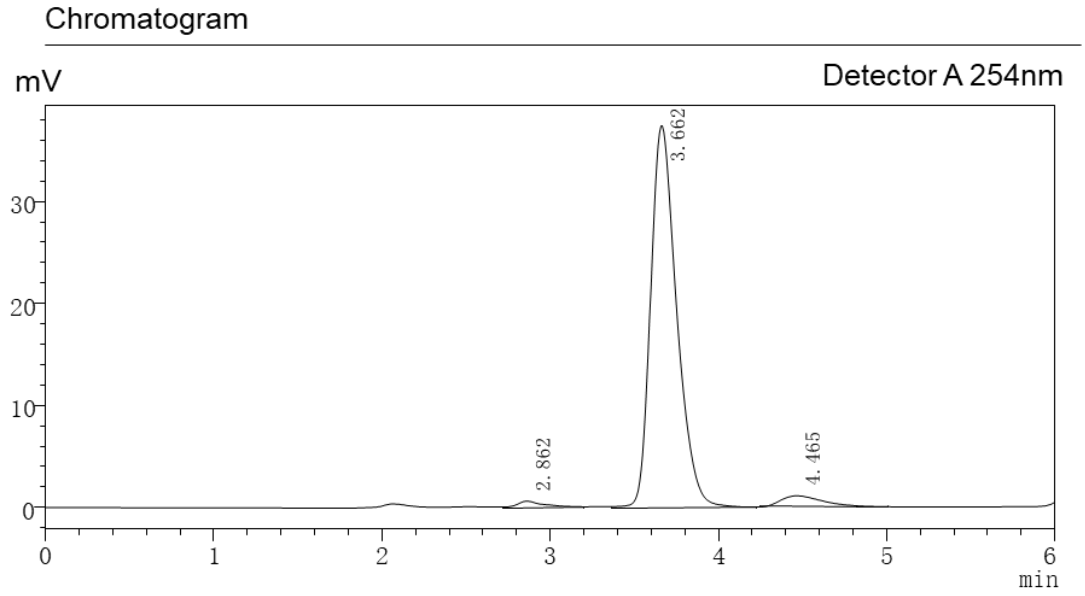


Supplementary Fig. 27 | ^{13}C -NMR spectrum of compound CN1. ^{13}C NMR (101 MHz, $CDCl_3$) δ 140.49, 139.89, 134.00, 129.48, 129.24, 127.96, 111.01, 105.67, 50.72.



Peak information (PDA Ch1 254nm)			
Peak	Retention time	Peak area	Peak area %
1	2.061	4011	0.989
2	4.456	394000	97.130
3	6.975	3045	0.751
4	14.611	4587	1.131
In total		405643	100.000

Supplementary Fig. 28 | HPLC chromatogram of compound CN1. The HPLC purity is 97.130%.



Peak information (PDA Ch1 254nm)			
Peak	Retention time	Peak area	Peak area %
1	2.862	7563	1.775
2	3.662	401436	94.192
3	4.465	17192	4.034
In total		426191	100.000

Supplementary Fig. 29 | HPLC chromatogram of compound CN2. The HPLC purity is 94.192%.

Reference

1. Dai, Q. et al. AncPhore: A versatile tool for anchor pharmacophore steered drug discovery with applications in discovery of new inhibitors targeting metallo-β-lactamases and indoleamine/tryptophan 2,3-dioxygenases. *Acta Pharm. Sin. B* **11**, 1931-1946 (2021).
2. Luo, S. et al. A 3D Generative Model for Structure-Based Drug Design. Preprint at https://arxiv.org/abs/2203.10446 (2022).
3. Peng, X. et al. Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets. Preprint at https://arxiv.org/abs/2205.07249 (2022).
4. Zhung, W., Kim, H. & Kim, W. Y. 3D molecular generative framework for interaction-guided drug design. *Nat. Commun.* 15, 2688, (2024).
5. Guan, J. et al. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. Preprint at https://arxiv.org/abs/2303.03543 (2023).
6. Schneuing, A. et al. Structure-based drug design with equivariant diffusion models. *Nat. Comput. Sci.* 4, 899-909 (2024).
7. Huang, L. et al. A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat. Commun.* 15, 2657, (2024).
8. Schneuing, A. et al. Multi-domain Distribution Learning for De Novo Drug Design. Preprint at https://openreview.net/forum?id=g3VCIM94ke (2025).
9. Adams, K. et al. ShEPhERD: Diffusing shape, electrostatics, and pharmacophores for bioisosteric drug design. Preprint at https://arxiv.org/abs/2411.04130 (2024).
10. Zhu, H., Zhou, R., Cao, D., Tang, J. & Li, M. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nat. Commun.* 14, 6234 (2023).
11. Francoeur, P. G. et al. Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design. *J. Chem. Inf. Model.* **60**, 4200-4215 (2020).
12. Hecker, S. J. et al. Discovery of Cyclic Boronic Acid QPX7728, an Ultrabroad-Spectrum Inhibitor of Serine and Metallo-β-lactamases. *J. Med. Chem.* **63**, 7491-7507 (2020).