Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2021

Supporting Information

Deep Generative Design with 3D Pharmacophoric Constraints

Fergus Imrie,[†] Thomas E. Hadfield,[†] Anthony R. Bradley,[‡] and Charlotte M. Deane[*,†]

†*Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, UK*

‡*Exscientia Ltd, The Schrödinger Building, Oxford Science Park, Oxford OX4 4GE, UK*

E-mail: deane@stats.ox.ac.uk

Additional DEVELOP model details

Atom types.

In line with both Liu et al.[1] and Imrie et al.[2], 14 atom types are permitted: carbon, nitrogen (N^-, N, N^+), oxygen (O^-, O, O^+), fluorine, chlorine, bromine, iodine, and sulphur (maximum valence 2, 4, or 6).

Hyperparameters.

As discussed in Methods, we used the same hyperparameters to train DEVELOP as adopted in Imrie et al.[2]. In particular, we trained the model with a learning rate of 0.001 for 10 epochs using the Adam optimiser and a batch size of 16. The hidden state dimension was set at 32, the encoding dimension 4, and λ_{KL} 0.3. The same hyperparameters were used for all experiments.

Comparison to SyntaLinker

SyntaLinker[3] is a transformer-based model for linker design that utilises a SMILES-based representation. The results in Table S4 are not directly comparable with the original publication since SyntaLinker was both trained and evaluated on different data sets in Yang et al.[3]. As described in Methods, we ensured a fair comparison between all methods by training and assessing all methods on the same data sets.

SyntaLinker produced substantially weaker results than were reported in its original publication.[3] In particular, SyntaLinker produced a very low proportion of valid molecules on both the CASF (Table S4) and PDBbind (Table 2) test sets. The poor validity of generated molecules can be explained by the sampling method used by SyntaLinker, which employs beam search[4] to generate SMILES strings. When sampling only the most likely sequences, the validity of generated molecules is relatively high; when the number of sequences sampled

is increased, the validity falls significantly as lower probability sequences are selected. In addition, we note that this sampling procedure limits the number of molecules that can be generated by SyntaLinker. However, as a result of using beam search, almost all of the SMILES strings generated by SyntaLinker correspond to unique molecules.

In addition to the low proportion of valid molecules generated, SyntaLinker recovered only 8% of the original molecules on the CASF set compared to 30% for DeLinker and 50% for DEVELOP, while on the PDBbind set SyntaLinker recovered 0.3% of the original molecules compared to 1.9% and 22.4% for DeLinker and DEVELOP, respectively. The 3D shape similarity of the molecules generated by SyntaLinker was significantly lower than DEVELOP for both the CASF (Table S4) and PDBbind (Table 2) test sets.

Differences in experimental setup. There are a number of differences between the experimental setup in this work and Yang et al.[3] which could be contributing to the differences in reported results for SyntaLinker.

First, Yang et al.[3] used a different filtering pipeline to construct their training and test sets. In particular, they applied Lipinski's "Rules of Five", filtered out pan assay interference compounds (PAINS) substructures,[5] and set an absolute cut-off for synthetic accessibility score[6] of 6.5. Then, they employed a fragmentation algorithm from matched-molecular pair analysis proposed by Hussain and Rea[7] that performs double cuts of non-functional group, acyclic single bonds. These candidate examples were then filtered for "Rule of three" criteria,[8] a maximum shortest linker bond distance of 15, a maximum synthetic accessibility score of each of the two starting substructures of 5, and a restriction that the synthetic accessibility score of the linker is lower than the sum of the two fragments. While many of these criteria are closely related to our data construction process, we note that they lead to different training and test sets for the same input data.

Second, Yang et al.[3] derived their training data from the ChEMBL[9] database, while we used the c. 250,000 compound subset of ZINC[10] selected at random by Gómez-Bombarelli et al.[11]. As a result, we note that the results reported in Yang et al.[3] comparing SyntaLinker

and DeLinker are not directly comparable due to the different training and test sets employed.

Assessment

Metrics

In this section, we provide additional details for the performance metrics used to assess the generative models

Validity. A molecule is deemed "valid" if it contains the starting substructure(s) and its SMILES representation can be parsed by RDKit[12] (i.e., satisfies atomic valency rules).

$$\text{Valid} = \frac{\text{\# Chemically valid SMILES strings}}{\text{\# Generated molecules}}$$

Uniqueness. Uniqueness measures the proportion of distinct molecules generated. Uniqueness was checked on a per-example basis to remove any dependency between examples in the test set. The total number of distinct molecules across all examples is divided by the number of valid generated molecules to calculate the proportion of unique molecules.

$$\text{Unique} = \frac{\text{\# Distinct valid molecules}}{\text{\# Valid molecules}}$$

Internal Diversity. Internal diversity measures the chemical diversity within a collection of molecules.[13,14] For a set of molecules, G, the internal diversity is given by:

$$\text{IntDiv}_p = 1 - \left(\frac{1}{|G|^2} \sum_{m_1, m_2 \in G} \text{Tanimoto Similarity}(m_1, m_2)^p \right)^{1/p}$$

A higher value corresponds to greater diversity in the set of molecules. Following Polykovskiy et al.[14], we report IntDiv_1 and IntDiv_2. IntDiv ranges from 0 to 1. Note this metric is calculated on a set of unique molecules (i.e. we filter duplicates before calculating internal diversity) to isolate molecular diversity from uniqueness.

Novelty. Novelty assesses the proportion of generated linkers or elaborations that were not present in the training set. The total number of molecules with novel elaborations is divided by the number of valid generated molecules to calculate the proportion of novel molecules.

$$\text{Novel} = \frac{\text{\# Valid elaborations not in training set}}{\text{\# Valid molecules}}$$

Additional figures

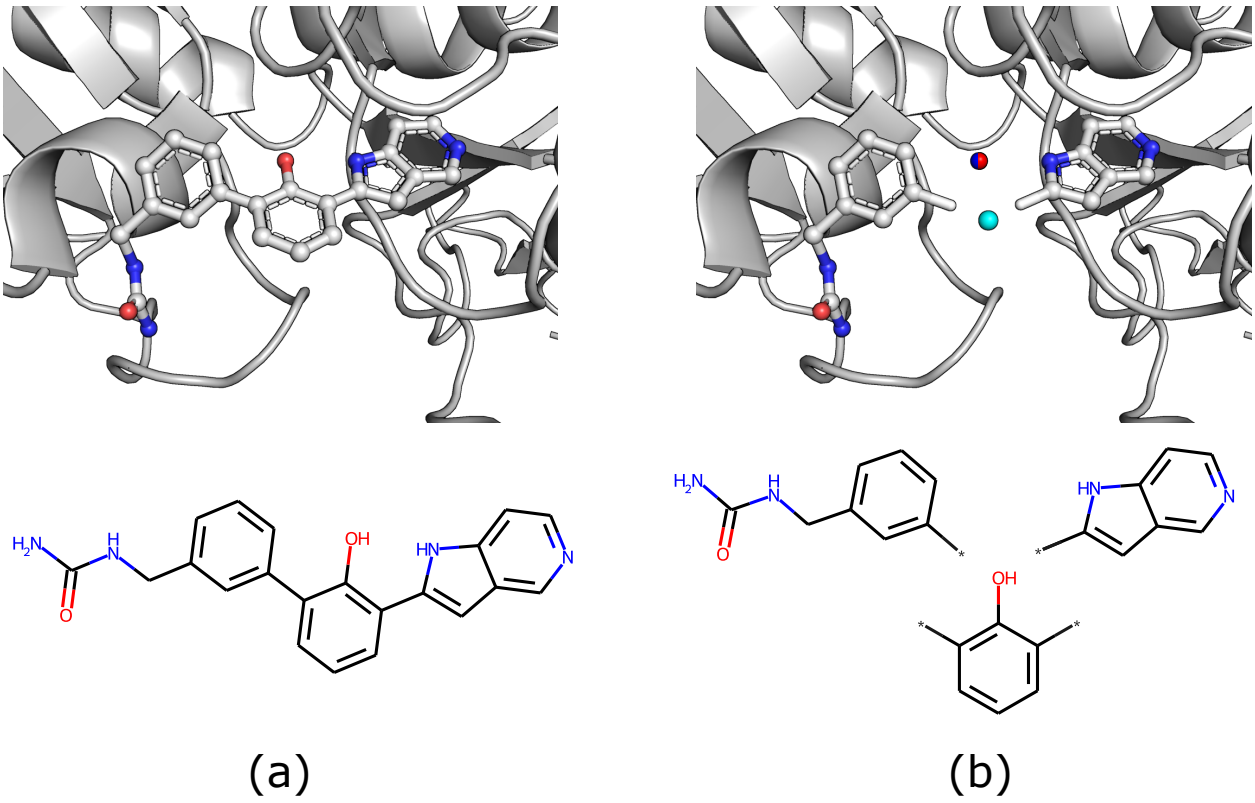

Figure S1: Example pharmacophoric featurisation for PDB ID 2FLR. (a) Top: Input PDB structure; Bottom: 2D depiction of ligand. (b) Top: Starting substructures and 3D pharmacophoric representation (cyan: aromatic pharmacophore, red: acceptor, blue: donor); Bottom: 2D depictions of starting substructures and target elaboration.

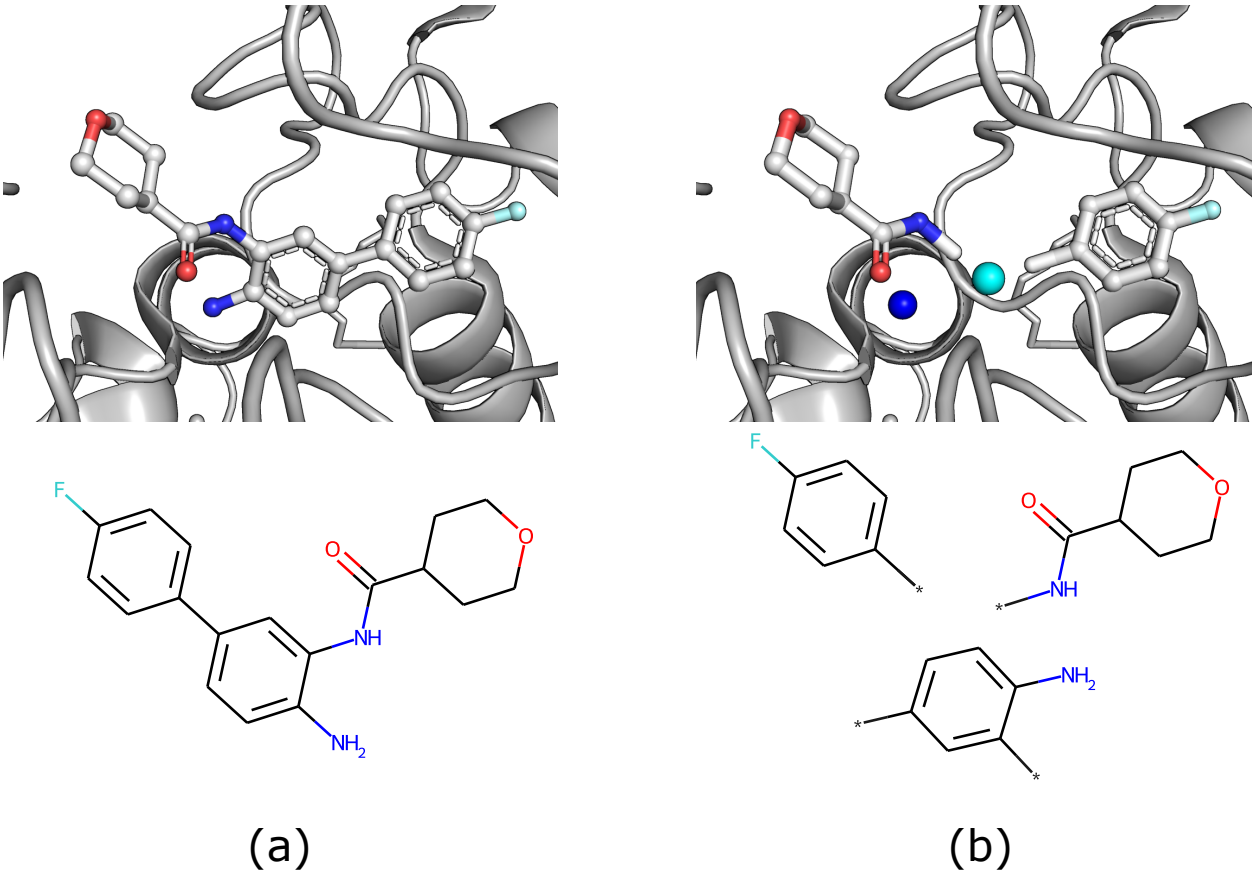

Figure S2: Example pharmacophoric featurisation for PDB ID 5IWG. (a) Top: Input PDB structure; Bottom: 2D depiction of ligand. (b) Top: Starting substructures and 3D pharmacophoric representation (cyan: aromatic pharmacophore, red: acceptor, blue: donor); Bottom: 2D depictions of starting substructures and target elaboration.

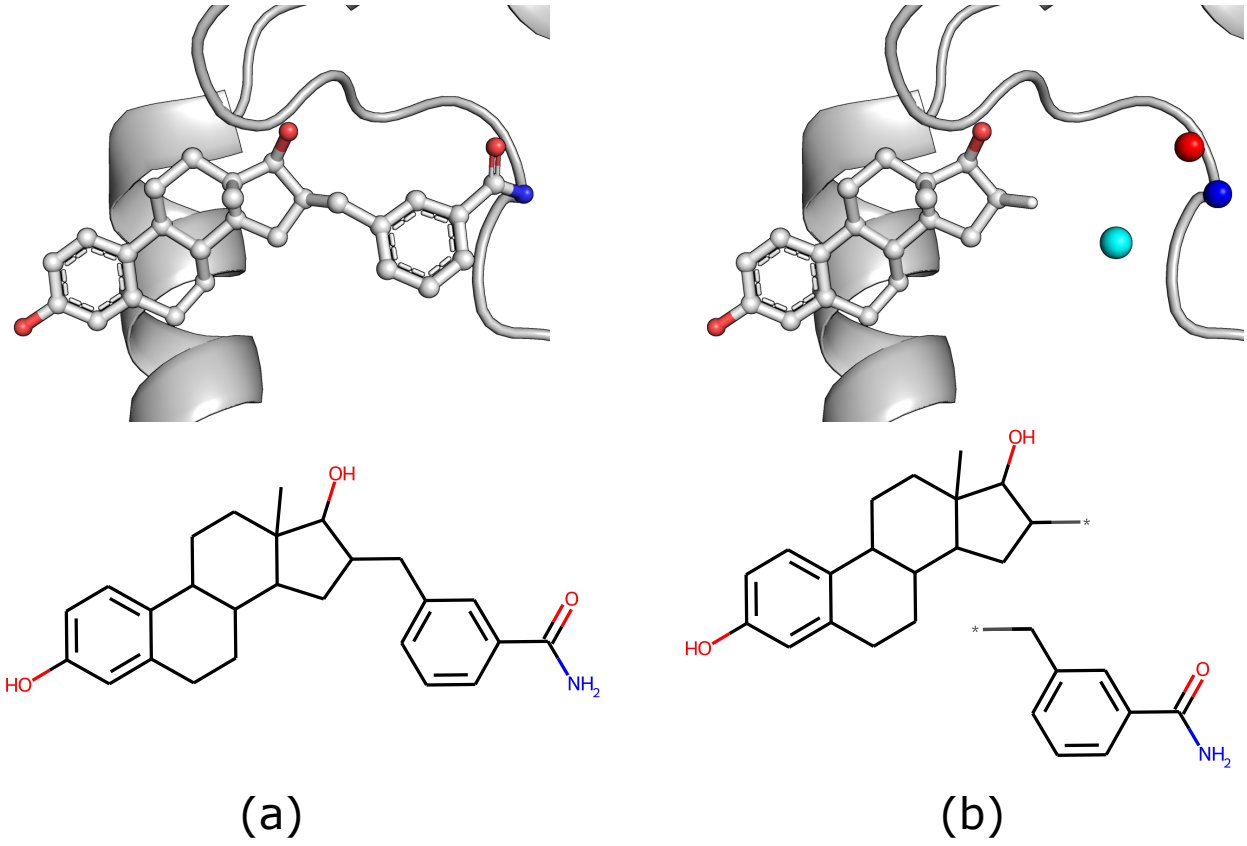

Figure S3: Example pharmacophoric featurisation for PDB ID 3HB4. (a) Top: Input PDB structure; Bottom: 2D depiction of ligand. (b) Top: Starting substructures and 3D pharmacophoric representation (cyan: aromatic pharmacophore, red: acceptor, blue: donor); Bottom: 2D depictions of starting substructures and target elaboration.

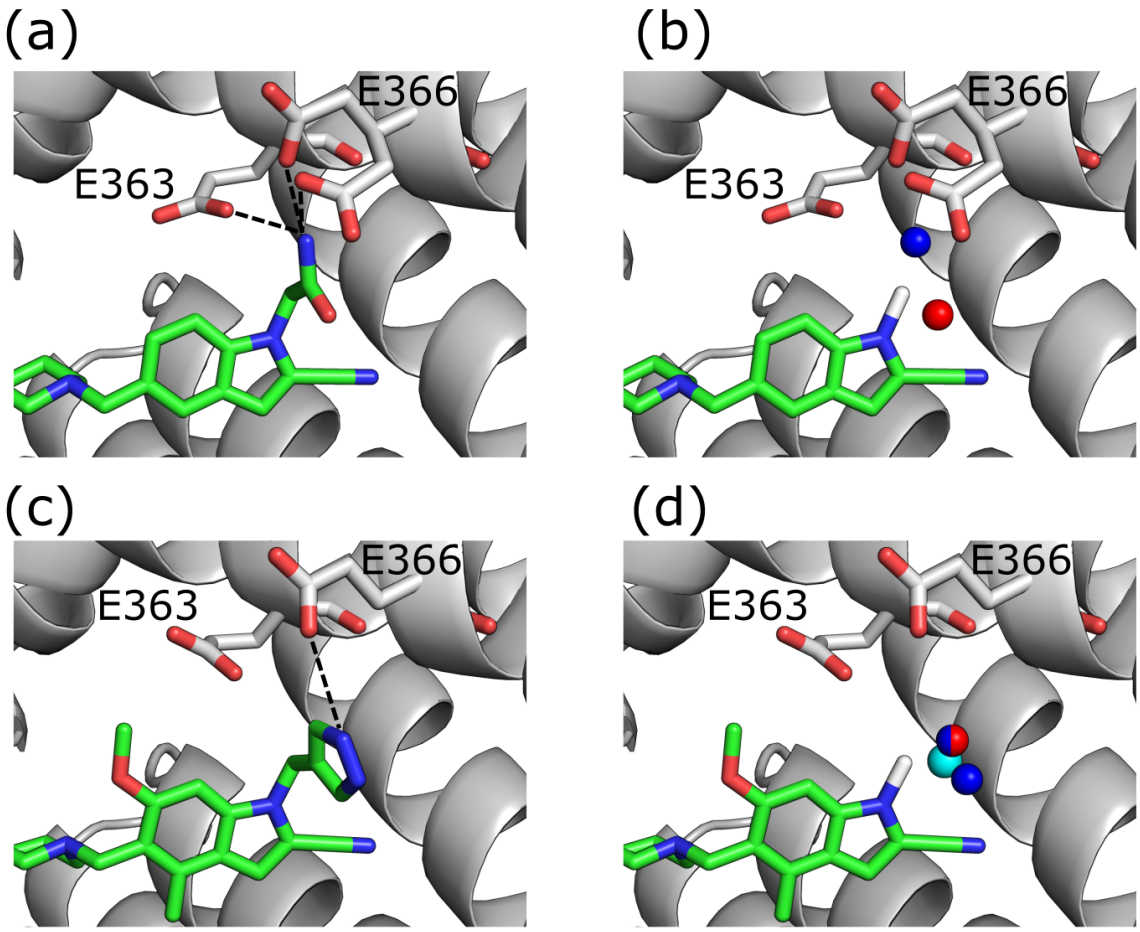

Figure S4: R-group optimisation case study pharmacophoric featurisation. Crystal structures 5DB2 (a) and 5DB3 (c) were used to derive target pharmacophoric profiles, shown in (b) and (d), respectively. Aromatic pharmacophores shown in cyan, hydrogen bond acceptors in red, and hydrogen bond donors in blue.

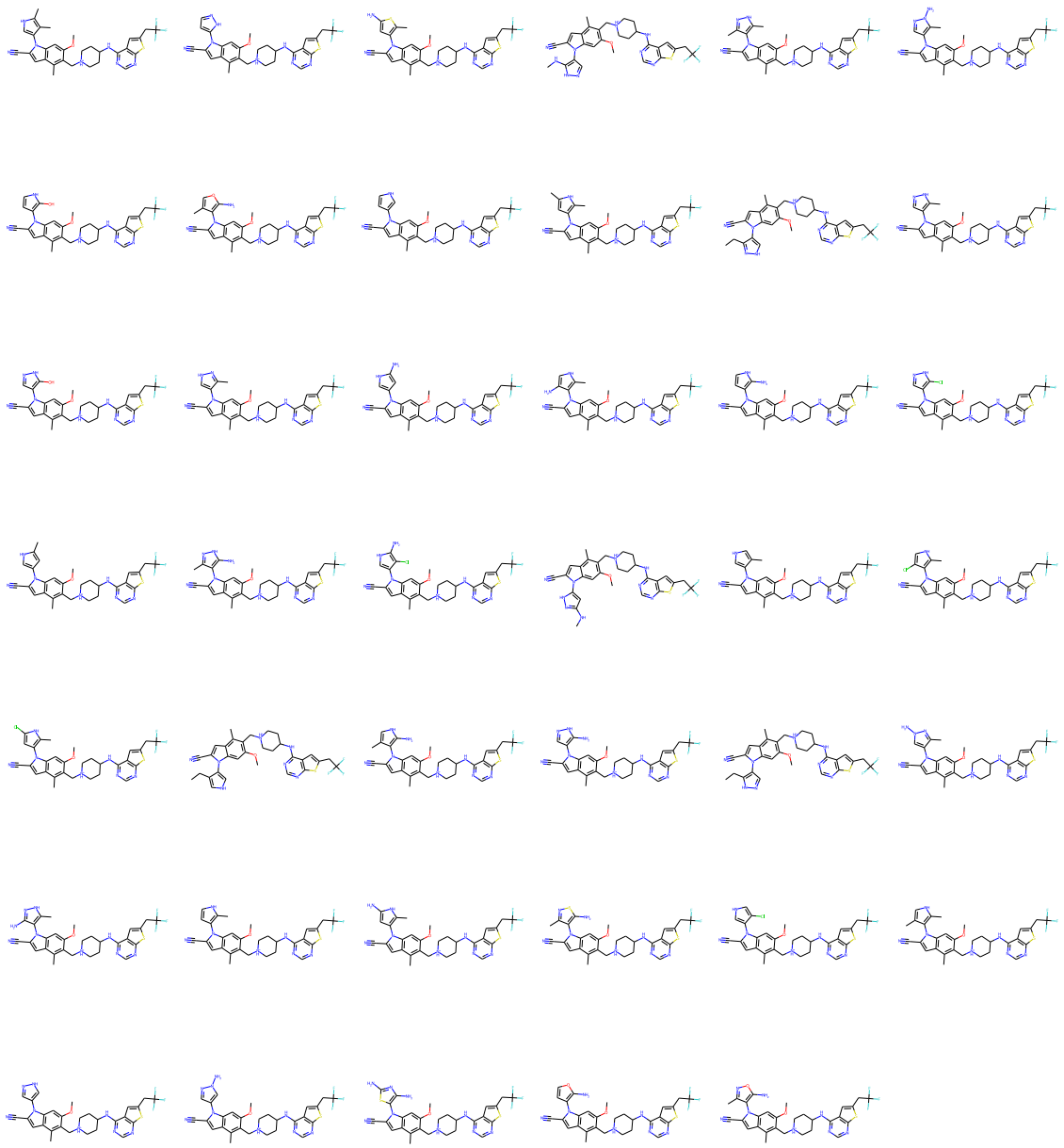

Figure S5: R-group optimisation case study. All 41 unique compounds generated by DEVELOP containing a hydrogen bond donor and aromatic ring for the Murcko scaffold shown in Figure 6c (left).

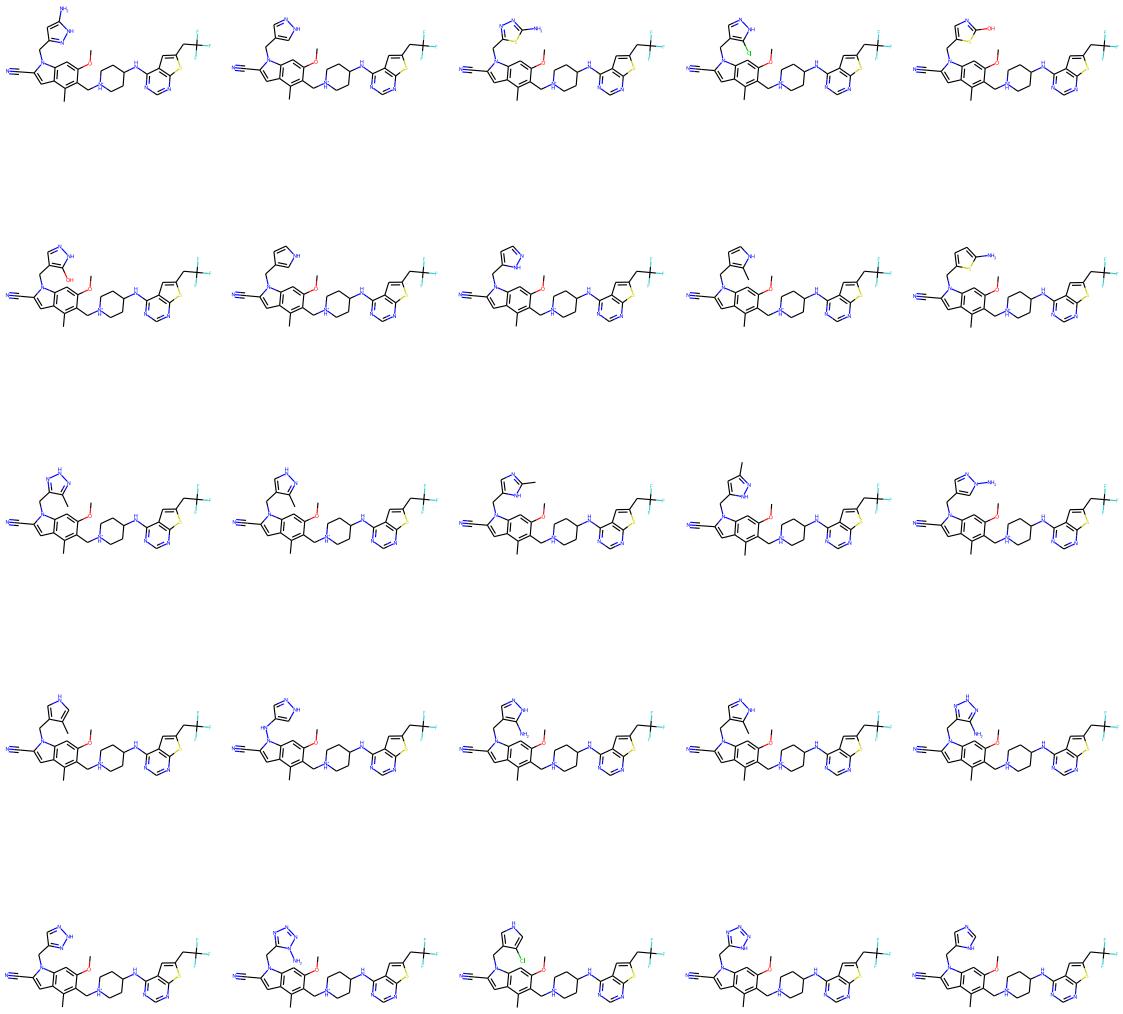

Figure S6: R-group optimisation case study. All 25 unique compounds generated by DEVELOP containing a hydrogen bond donor and aromatic ring for the Murcko scaffold shown in Figure 6c (middle).

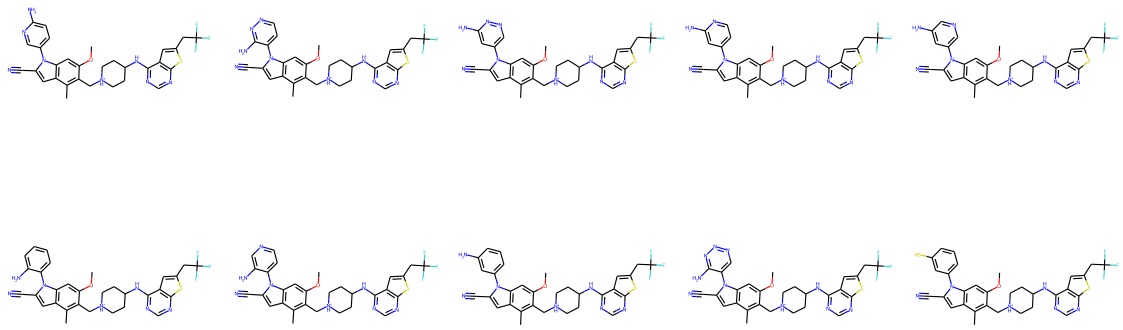

Figure S7: R-group optimisation case study. All 10 unique compounds generated by DEVELOP containing a hydrogen bond donor and aromatic ring for the Murcko scaffold shown in Figure 6c (right).

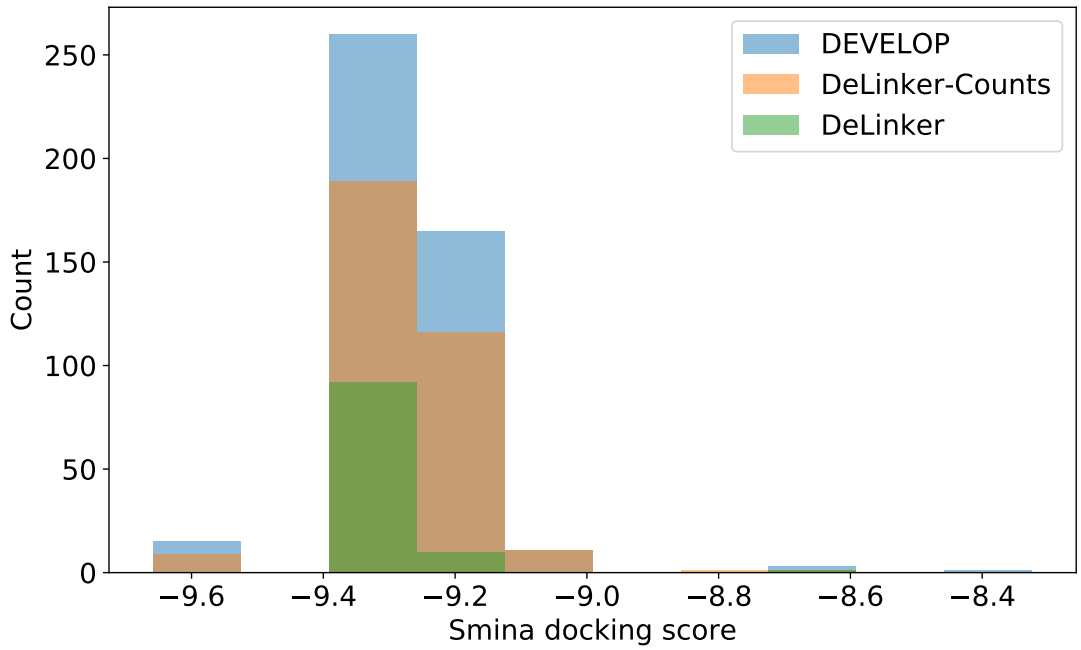

Figure S8: R-group optimisation case study. Predicted binding affinities from docking using the smina[15] version of AutoDock Vina[16] for molecules that satisfied the pharmacophoric profile of the acetamide group (Figure 6b, left).

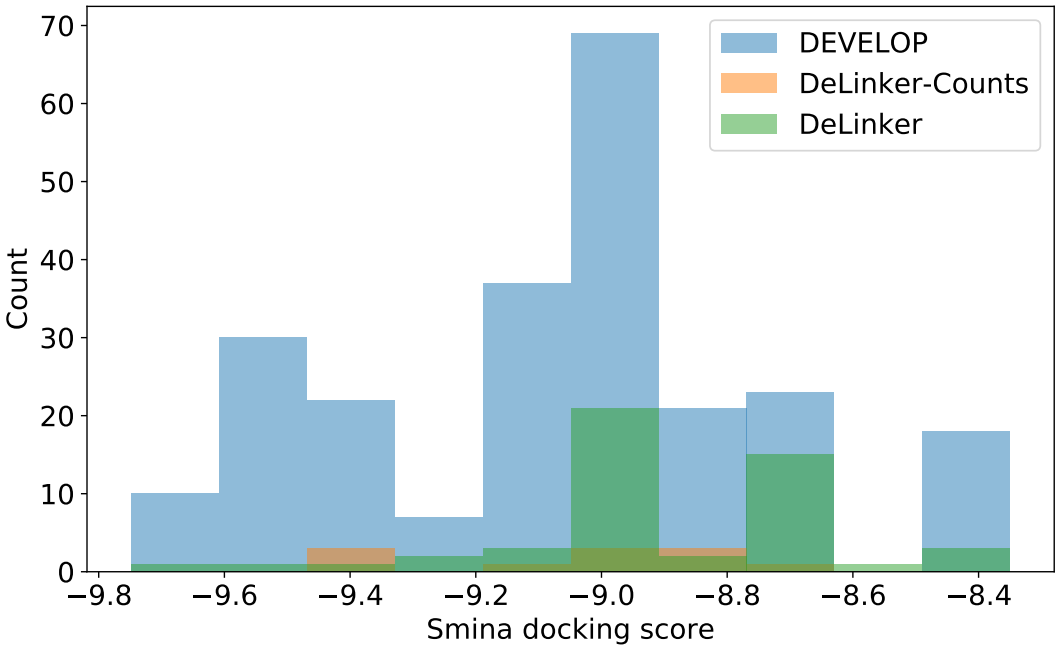

Figure S9: R-group optimisation case study. Predicted binding affinities from docking using the smina[15] version of AutoDock Vina[16] for molecules that satisfied the pharmacophoric profile of the 4-methylpyrazole elaboration (Figure 6b, right).

Additional results

Table S1: Linker design. Structural diversity metrics.

Metric	SyntaLinker	DeLinker	DeLinker-Counts	DEVELOP
CASF				
Unique	8.1%	72.9%	58.6%	58.2%
$IntDiv_1$	0.532	0.563	0.546	0.541
$IntDiv_2$	0.524	0.560	0.542	0.536
PDBbind				
Unique	93.6%	72.9%	58.6%	58.2%
$IntDiv_1$	0.553	0.607	0.587	0.576
$IntDiv_2$	0.546	0.604	0.583	0.572

Table S2: Scaffold elaboration. Structural diversity metrics.

Metric	Scaffold-Decorator	DeLinker	DeLinker-Counts	DEVELOP
CASF				
Unique	25.2%	74.2%	52.0%	39.7%
$IntDiv_1$	0.440	0.468	0.449	0.425
$IntDiv_2$	0.436	0.465	0.445	0.420
PDBbind				
Unique	23.4%	87.8%	81.6%	76.2%
$IntDiv_1$	0.429	0.491	0.472	0.457
$IntDiv_2$	0.426	0.488	0.468	0.453

Table S3: Linker design. PDBbind set results comparing the two different sampling methods for SyntaLinker (beam search and random sampling).

Metric	SyntaLinker Beam Search	SyntaLinker Sampling	DeLinker	DeLinker-Counts	DEVELOP
Valid	11.6%	65.4%	96.9%	90.2%	93.1%
Unique	93.6%	9.0%	86.1%	77.8%	77.3%
Novel	54.8%	9.9%	84.0%	87.6%	88.7%
Recovered	0.3%	0.0%	1.9%	8.7%	22.4%
Pass 2D filters	81.1%	95.1%	63.4%	59.5%	61.7%
SC_{RDKit} Generated					
>0.6	9.4%	13.4%	10.4%	19.8%	27.9%
>0.7	4.5%	7.7%	4.2%	10.1%	14.8%
>0.8	2.2%	4.8%	1.5%	4.4%	6.1%
>0.9	0.7%	1.3%	0.4%	1.2%	1.5%

Table S4: Linker design. CASF set results, including two different sampling methods for SyntaLinker (beam search and random sampling).

Metric	SyntaLinker Beam Search	SyntaLinker Sampling	DeLinker	DeLinker-Counts	DEVELOP
Valid	9.8%	58.3%	94.7%	86.0%	89.6%
Unique	95.1%	8.1%	72.9%	58.6%	58.2%
Novel	54.4%	13.1%	68.7%	68.4%	71.1%
Recovered	7.5%	5.9%	29.8%	41.5%	50.0%
Pass 2D filters	80.5%	92.8%	71.7%	71.7%	68.6%
SC_{RDKit} Generated					
>0.6	14.0%	27.4%	23.0%	40.6%	45.5%
>0.7	7.0%	18.1%	12.2%	27.9%	31.0%
>0.8	3.8%	12.2%	6.4%	18.4%	20.7%
>0.9	1.5%	6.6%	2.7%	10.3%	12.2%

Table S5: Linker design. Alternative 3D similarity metrics for CASF test set.

Metric	SyntaLinker	DeLinker	DeLinker-Counts	DEVELOP
SC_{RDKit} Molecule				
>0.7	22.5%	16.3%	22.5%	25.5%
>0.8	4.7%	3.6%	7.0%	7.2%
>0.9	0.5%	0.8%	2.9%	3.0%
SC_{RDKit} Fragments				
>0.7	42.6%	38.7%	40.8%	43.0%
>0.8	14.2%	12.3%	14.7%	15.4%
>0.9	1.3%	1.6%	3.2%	3.8%
RMSD				
<1.00Å	30.5%	26.6%	28.1%	30.7%
<0.75Å	11.8%	9.3%	10.6%	12.9%
<0.50Å	3.0%	2.4%	4.3%	4.7%

Table S6: Linker Design. Alternative 3D similarity metrics for PDBbind test set.

Metric	SyntaLinker	DeLinker	DeLinker-Counts	DEVELOP
SC_{RDKit} Molecule				
>0.7	12.6%	12.8%	15.8%	17.3%
>0.8	2.4%	1.9%	3.3%	3.7%
>0.9	0.1%	0.0%	0.2%	0.2%
SC_{RDKit} Fragments				
>0.7	35.0%	34.5%	35.7%	34.0%
>0.8	11.8%	9.9%	10.2%	10.3%
>0.9	2.0%	1.1%	0.8%	1.0%
RMSD				
<1.00Å	26.7%	24.8%	26.1%	25.0%
<0.75Å	10.3%	8.5%	8.9%	8.7%
<0.50Å	3.0%	1.4%	1.4%	1.6%

Table S7: Scaffold elaboration. CASF set results.

Metric	Scaffold-Decorator	DeLinker	DeLinker-Counts	DEVELOP
Valid	99.9%	100.0%	100.0%	99.8%
Unique	25.2%	74.2%	52.0%	39.7%
Novel	2.4%	55.1%	50.7%	43.4%
Recovered	18.9%	33.6%	45.5%	58.7%
Pass 2D filters	98.4%	64.2%	67.5%	71.7%
SC_{RDKit} Generated				
>0.6	18.6%	14.1%	32.3%	51.2%
>0.7	10.6%	7.0%	23.2%	37.7%
>0.8	5.6%	2.7%	15.1%	24.8%
>0.9	1.3%	1.3%	6.2%	10.4%

Table S8: Scaffold elaboration. PDBbind set results, SC_{RDKit} Generated calculated with RDKit generated reference conformers.

Metric	Scaffold-Decorator	DeLinker	DeLinker-Counts	DEVELOP
Valid	99.9%	100.0%	100.0%	99.5%
Unique	23.4%	87.8%	81.6%	76.2%
Novel	2.0%	71.1%	79.2%	78.2%
Recovered	0.0%	1.0%	4.5%	15.3%
Pass 2D filters	98.9%	55.3%	47.8%	51.3%
SC_{RDKit} Generated				
>0.6	10.3%	7.6%	13.0%	31.5%
>0.7	4.3%	2.7%	5.7%	16.9%
>0.8	0.6%	0.7%	1.8%	6.8%
>0.9	0.0%	0.1%	0.3%	1.5%

Table S9: Scaffold elaboration. CASF set results, SC_{RDKit} Generated calculated with RDKit generated reference conformers.

Metric	Scaffold-Decorator	DeLinker	DeLinker-Counts	DEVELOP
Valid	99.9%	100.0%	100.0%	99.8%
Unique	25.2%	74.2%	52.0%	39.7%
Novel	2.4%	55.1%	50.7%	43.4%
Recovered	18.9%	33.6%	45.5%	58.7%
Pass 2D filters	98.4%	64.2%	67.5%	71.7%
SC_{RDKit} Generated				
>0.6	30.3%	22.4%	42.8%	67.6%
>0.7	22.0%	13.9%	35.8%	58.6%
>0.8	12.0%	5.3%	24.8%	44.0%
>0.9	5.1%	3.1%	18.1%	33.7%

References

(1) Liu, Q.; Allamanis, M.; Brockschmidt, M.; Gaunt, A. Constrained Graph Variational Autoencoders for Molecule Design. *Advances in Neural Information Processing Systems 31 (NeurIPS)* **2018**, 7795–7804.

(2) Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M. Deep Generative Models for 3D Linker Design. *J. Chem. Inf. Model.* **2020**, *60*, 1983–1995.

(3) Yang, Y.; Zheng, S.; Su, S.; Zhao, C.; Xu, J.; Chen, H. SyntaLinker: automatic fragment linking with deep conditional transformer neural networks. *Chem. Sci.* **2020**, *11*, 8312–8322.

(4) Lowerre, B. T. The Harpy Speech Recognition System. Ph.D. thesis, 1976.

(5) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem* **2010**, *53*, 2719–2740.

(6) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminf.* **2009**, *1*, 8.

(7) Hussain, J.; Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 339–348.

(8) Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discovery* **2013**, *12*, 644–644.

(9) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **2011**, *40*, D1100–D1107.

(10) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337.

(11) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.

(12) Landrum, G. RDKit: Open-Source Cheminformatics. `http://www.rdkit.org/`, (accessed November 4, 2019).

(13) Benhenda, M. ChemGAN challenge for drug discovery: Can AI reproduce natural chemical diversity? *arXiv preprint arXiv:1708.08227* **2017**,

(14) Polykovskiy, D. et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2020**, *11*, 1931.

(15) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with smina From the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904.

(16) Trott, O.; Olson, A. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.