Supporting Information

Deep Generative Models for 3D Linker Design

Fergus Imrie,[†] Anthony R. Bradley,[‡] Mihaela van der Schaar,[¶,§] and Charlotte M. Deane[*,†]

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

‡*Exscientia Ltd, 36 St. Giles', Oxford, OX1 3LD, UK*

¶*University of Cambridge, Cambridge CB2 1PZ, UK*

§*Alan Turing Institute, London NW1 2DB, UK*

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

DeLinker Implementation Details

Atom types.

There are 14 permitted atom types: carbon, nitrogen (N^-, N, N^+), oxygen (O^-, O, O^+), fluorine, chlorine, bromine, iodine, and sulphur (maximum valence 2, 4, or 6).

Network architecture.

Following Liu et al.[1], both the encoder and decoder are standard gated graph neural networks (GGNN),[2] which propagate messages for 7 steps, and have residual connections between odd numbered time steps.

We implemented the function f, which maps the hidden state of a node to its atom type, as a linear classifier with attention from the node's hidden vector to one of the node types. The attention mechanism is similar to Bahdanau et al.[3] and allows the label for a given node to depend on the hidden states of the other expansion nodes.

Similarly, we augmented the edge selection and edge labelling step by adding attention between the feature vectors for all candidate edges. This allows the score of a candidate edge to depend on the other possible edges.

We trained the model with a learning rate of 0.001 for 10 epochs using the Adam optimiser.

Hyperparameter search.

We performed a limited hyperparameter search of the following parameters (final parameters in bold):

- Learning rate: 0.01, **0.001**, 0.0001
- Batch size: 8, **16**
- Hidden state dimension: 16, **32**, 50, 100

- Encoding dimension: 2, **4**, 8
- λ_{KL}: 0.1, **0.3**, 0.6

The model was fairly robust to the choice of hyperparameters. Performance was measured via the validation reconstruction loss and not generative performance.

Data curation

Fragment-molecule pairs for the ZINC[4] and CASF[5] sets were constructed as follows. First all possible fragmentations of each molecule were produced by enumerating all double acyclic single bond cuts.[6] These we then filtered to remove trivial and unrealistic situations using the following constraints: (i) minimum linker length: 3 atoms, (ii) minimum fragment size: 5 atoms, (iii) linker fewer heavy atoms than either fragment, (iv) minimum path length between fragments: 2 atoms.

The remaining fragment-molecule pairs were filtered for several 2D properties: (i) the synthetic accessibility (SA) score[7] of the molecule must be lower than the fragments with exit vectors represented by dummy atoms, (ii) the molecule must pass pan-assay interference (PAINS)[8] filters, and (iii) rings must either be saturated aliphatic or aromatic (according to RDKit[9] valency rules). PAINS filters were implemented by SMARTS substructure searching with RDKit, using the RDKit version of the Saubern et al.[10] translation of the original PAINS.[8] as specified at `https://github.com/rdkit/rdkit/blob/master/Data/Pains/wehi_pains.csv` (Accessed: 02/06/2019). Any molecules containing atom types outside of the permitted atom types were excluded.

Training set composition

Table S1: Distribution of number of atoms contained in the original linkers in the datasets utilised. The average linker length in the CASF set (5.9) is around one atom longer than the ZINC training set (4.7), validation (4.7) and test set (4.9).

Linker	ZINC			CASF
Length	Train	Valid	Test	Test
3	28.7%	30.7%	26.2%	25.2%
4	21.6%	22.7%	17.7%	13.9%
5	19.5%	15.7%	22.0%	10.7%
6	17.8%	16.5%	21.0%	13.3%
7	7.7%	10.3%	7.3%	12.0%
8	3.1%	2.5%	3.5%	9.1%
9	1.3%	1.3%	2.0%	4.8%
10	0.3%	0.3%	0.3%	3.2%
11	0.0%	-	-	5.5%
$\geq$12	0.0%	-	-	2.3%

Additional results

Table S2: Ablation study for DeLinker, our deep generative method on the ZINC data set. We show the effect on the 2D metrics of removing all of the structural information ("No info") and including only the distance information ("Distance") compared to our full protocol ("DeLinker"), the database baseline ("Database"), and a graph-based baseline[1]("CGVAE"). See Data curation for a description of the 2D property filters.

Metric	Database	CGVAE	No Info	Distance	DeLinker
Valid	100.0%	88.9%	97.0%	98.6%	98.4%
Unique	38.8%	58.8%	51.2%	47.3%	44.2%
Novel	0.0%	51.0%	36.2%	37.6%	39.5%
Recovered	78.0%	65.8%	74.5%	78.3%	79.0%
Pass 2D filters	97.0%	85.9%	89.9%	90.2%	89.8%
Pass SA filter	97.8%	90.0%	95.1%	95.5%	95.3%
Pass ring filter	100.0%	93.2%	95.2%	94.5%	94.8%
Pass PAINS filter	99.2%	96.1%	97.8%	98.4%	97.9%

Table S3: Ablation study for DeLinker, our deep generative method, on the ZINC data set. We show the effect on the 3D metrics of removing all of the structural information ("No info") and including only the distance information ("Distance") compared to our full protocol ("DeLinker") that includes both distance and angle information, the database baseline ("Database"), and a graph-based baseline[1]("CGVAE"). See Methods - Assessment metrics for a description of the metrics.

Metric	Database	CGVAE	No Info	Distance	DeLinker
SC_{RDKit} Molecule					
>0.7	35.5%	35.4%	37.6%	43.2%	47.1%
>0.8	8.5%	7.2%	9.2%	11.8%	14.2%
>0.9	1.3%	0.7%	1.1%	1.5%	1.8%
SC_{RDKit} Fragments					
>0.7	60.2%	64.1%	64.4%	69.1%	71.3%
>0.8	24.7%	26.3%	27.7%	33.4%	35.8%
>0.9	4.5%	4.2%	5.0%	7.0%	8.2%
RMSD Fragments					
<1.00	46.9%	51.0%	50.9%	56.6%	58.6%
<0.75	20.5%	21.6%	22.4%	27.8%	30.0%
<0.50	5.7%	4.8%	5.6%	7.9%	9.3%

Table S4: 2D and 3D metrics for molecules generated by DeLinker, our *de novo* deep generative model, compared to a Database baseline on the held-out ZINC test set. See Data curation for a description of the 2D property filters and Methods - Assessment metrics for a description of the 3D metrics.

	ZINC		ZINC $\geq$ 5 atoms	
Metric	Database	DeLinker	Database	DeLinker
Valid	100.0%	98.4%	100.0%	98.1%
Unique	38.8%	44.2%	53.6%	61.0%
Novel	0.0%	39.5%	0.0%	49.4%
Recovered	78.0%	79.0%	67.0%	67.0%
Pass 2D filters	97.0%	89.8%	96.4%	84.1%
SC_{RDKit} Molecule				
>0.7	33.5%	47.1%	21.3%	37.1%
>0.8	8.5%	14.2%	3.5%	9.4%
>0.9	1.3%	1.8%	0.4%	1.0%
SC_{RDKit} Fragments				
>0.7	60.2%	71.3%	51.5%	66.7%
>0.8	24.7%	35.8%	16.8%	30.3%
>0.9	4.5%	8.2%	2.1%	6.0%
RMSD Fragments				
<1.00Å	46.9%	58.6%	39.1%	55.1%
<0.75Å	20.5%	30.0%	14.2%	26.9%
<0.50Å	5.7%	9.3%	3.0%	6.9%

Table S5: Fragment linking case study. 2D and 3D Metrics for DeLinker and the Database baseline.

Metric	Database	DeLinker
Valid	100.0%	98.7%
Unique	30.7%	56.4%
Novel	0.0%	58.5%
Recovered	100.0%	100.0%
Pass 2D filters	97.8%	74.0%
SC_{RDKit} Fragments		
>0.7	681	1115
>0.8	129	301
>0.9	6	18

Table S6: Scaffold hopping case study. 2D and 3D metrics for DeLinker. Compounds with SC_{RDKit} Fragments>0.80 were docked with AutoDock Vina.[11,12]

Metric	DeLinker
Valid	99.5%
Unique	63.7%
Novel	88.5%
Recovered	100.0%
Pass 2D filters	51.4%
SC_{RDKit} Fragments	
>0.70	1928
>0.75	699
>0.80	114
>0.85	9
Vina Score	
<-7	105
<-8	69
<Aminopyrazole	33
<-9	26
<-10	3
<Indazole	0

Table S7: PROTAC design case study. 2D and 3D metrics for DeLinker. Compounds with SC_{RDKit} Fragments>0.80 were docked with AutoDock Vina[11,12]

Metric	DeLinker
Valid	96.0%
Unique	62.1%
Novel	95.9%
Recovered	0.0%
Pass 2D filters	61.8%
SC_{RDKit} Fragments	
>0.70	2930
>0.80	2930
>0.85	2150
Vina Score	
<-12	2930
<PROTAC 1	2927
<-13	2845
<-14	1160
<PROTAC 2	536
<-15	34

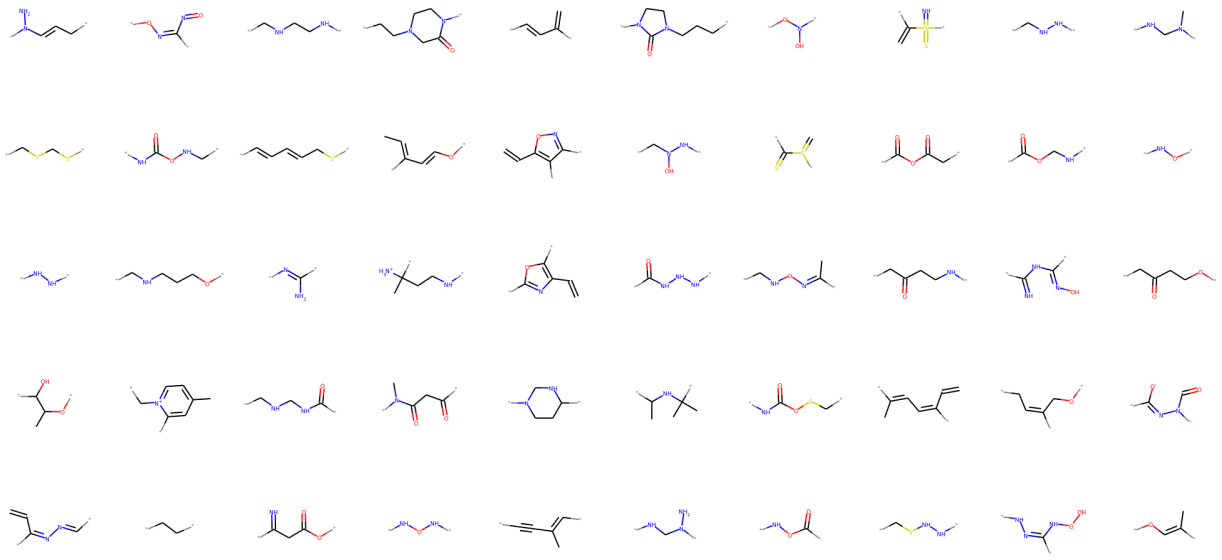

Figure S1: A random sample of 50 novel linkers generated by DeLinker during testing on the held-out ZINC data set.

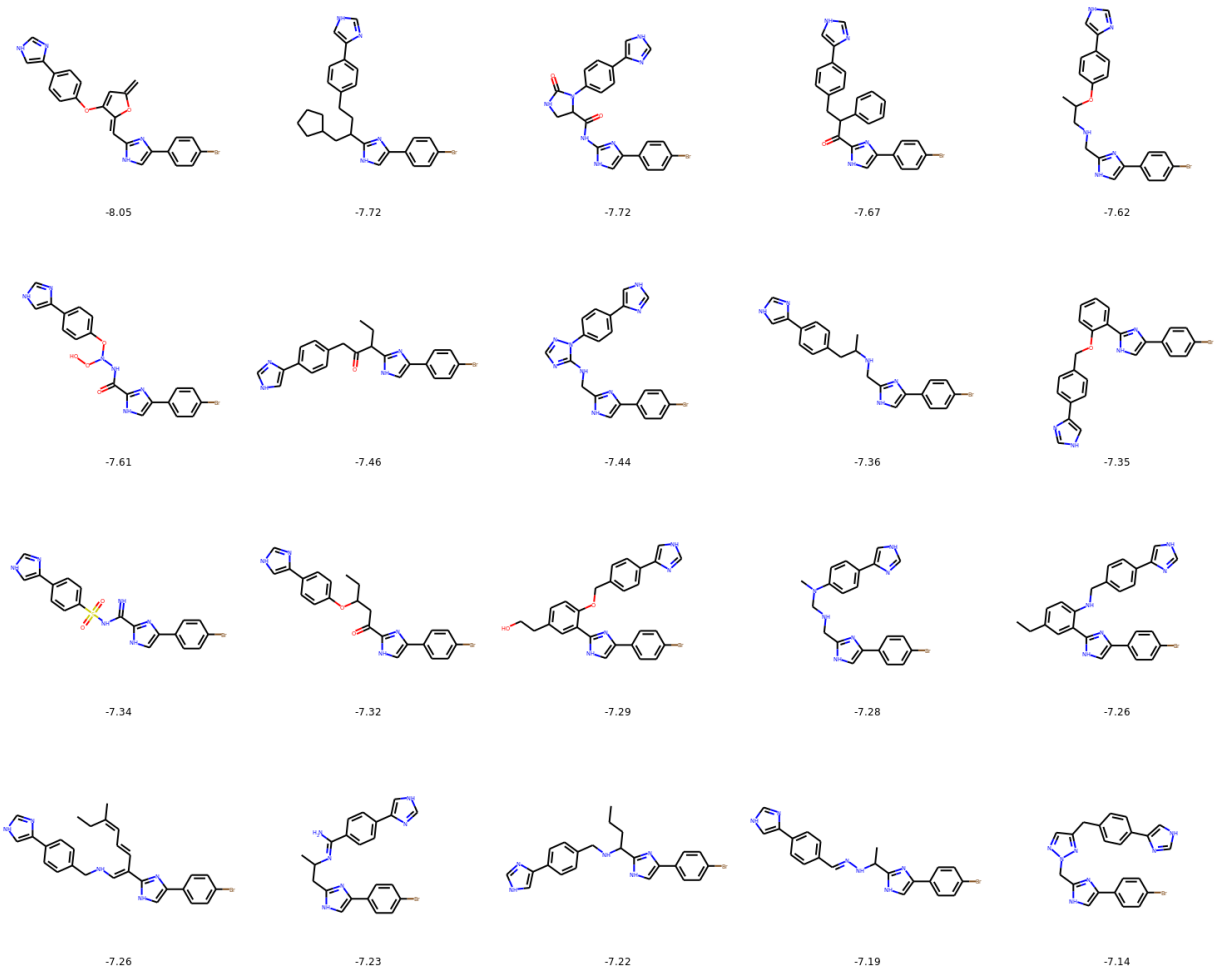

Figure S2: Fragment linking case study. The top 20 molecules generated by DeLinker that met the 3D similarity threshold ranked by AutoDock Vina[11,12] score. Labels are the docking score from minimizing the aligned molecular conformer according to the Vina energy function.

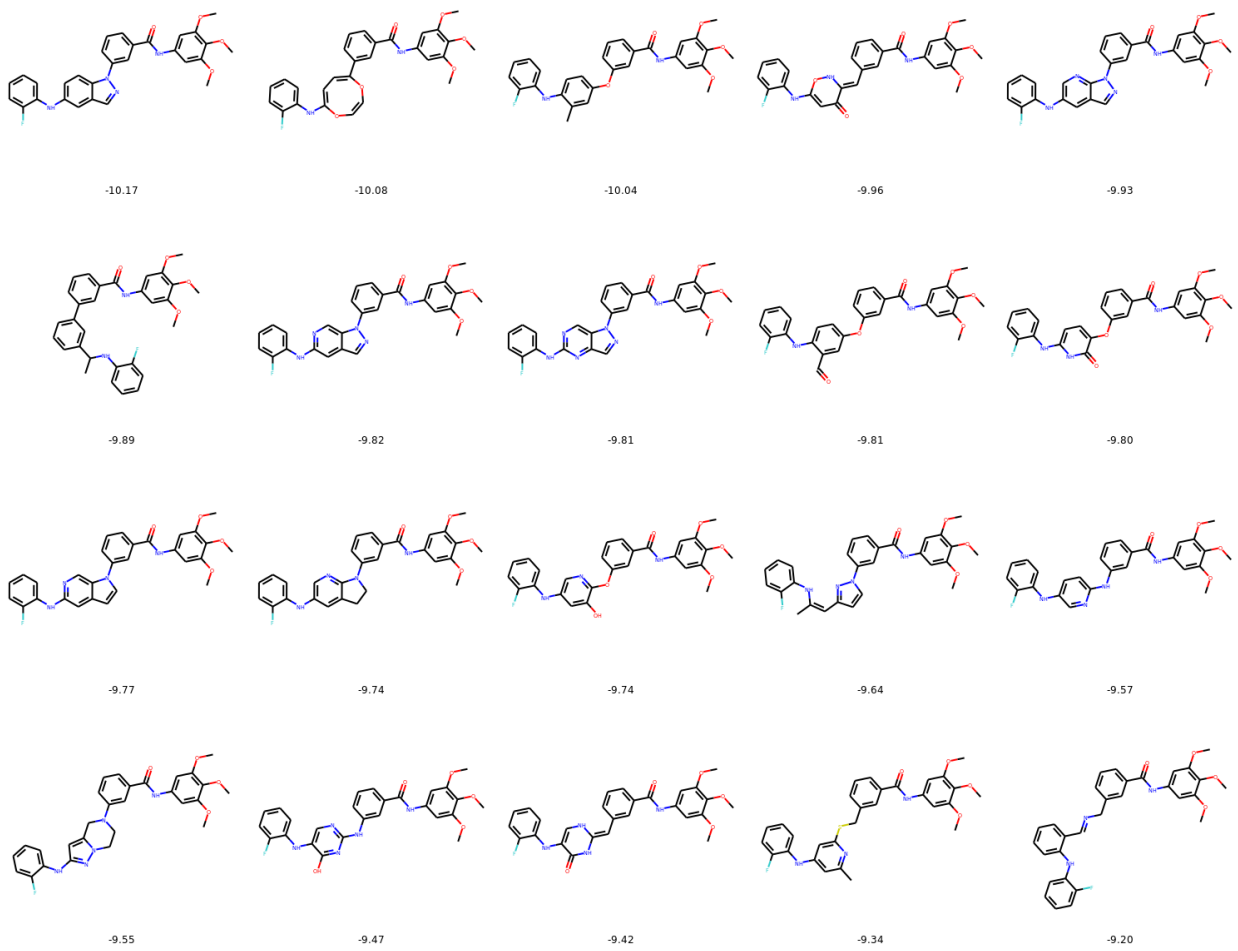

Figure S3: Scaffold hopping case study. The top 20 molecules generated by DeLinker that met the 3D similarity threshold ranked by AutoDock Vina[11,12] score. Labels are the docking score from minimizing the aligned molecular conformer according to the Vina energy function.

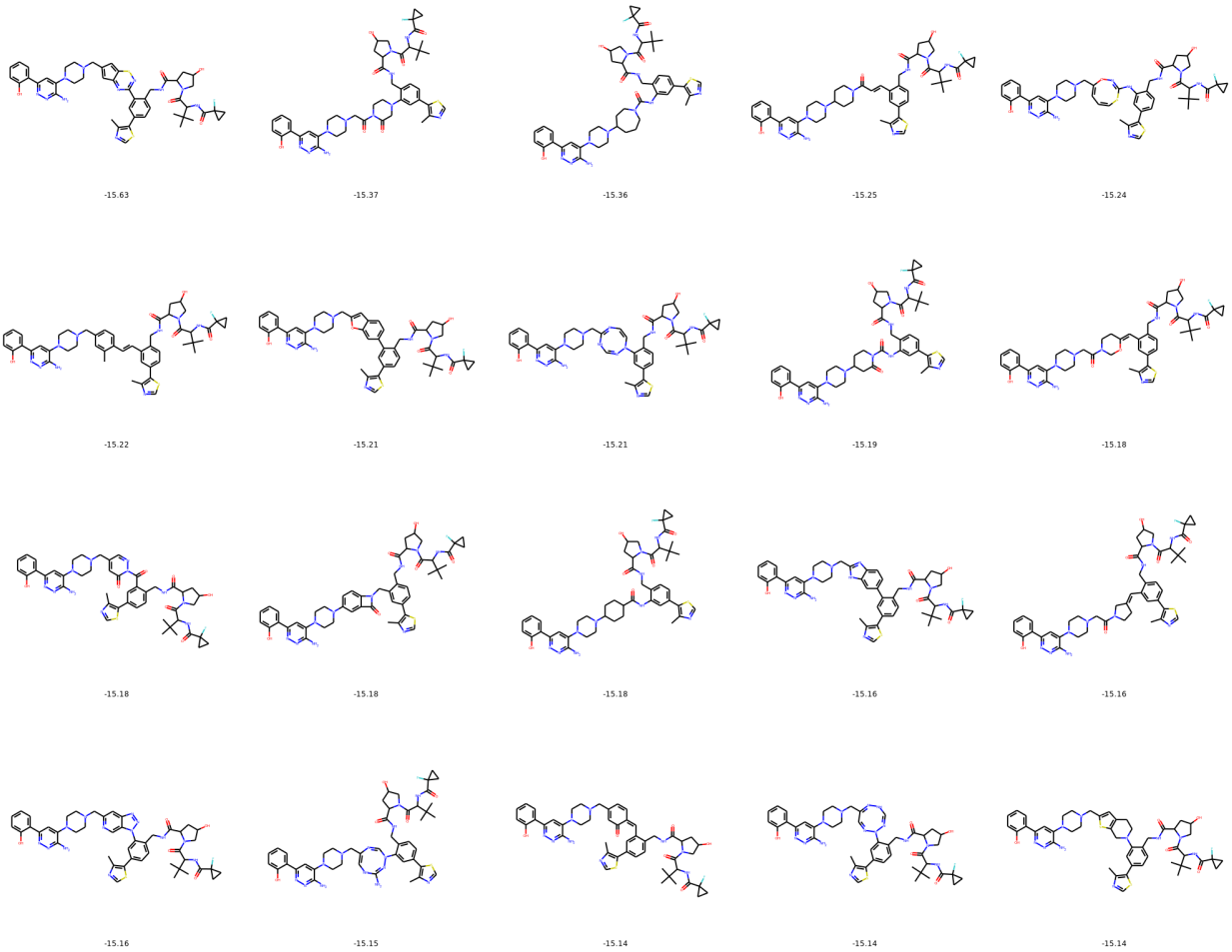

Figure S4: PROTAC design case study. The top 20 molecules generated by DeLinker that met the 3D similarity threshold ranked by AutoDock Vina[11,12] score. Labels are the docking score from minimizing the aligned molecular conformer according to the Vina energy function.

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) Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated Graph Sequence Neural Networks. *International Conference on Learning Representations (ICLR)* **2016**,

(3) Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations (ICLR)* **2015**,

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

(5) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59*, 895–913.

(6) 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.

(7) 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.

(8) 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.

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

(10) Saubern, S.; Guha, R.; Baell, J. B. KNIME Workflow to Assess PAINS Filters in SMARTS Format. Comparison of RDKit and Indigo Cheminformatics Libraries. *Mol. Inf.* **2011**, *30*, 847–850.

(11) 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.

(12) 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.