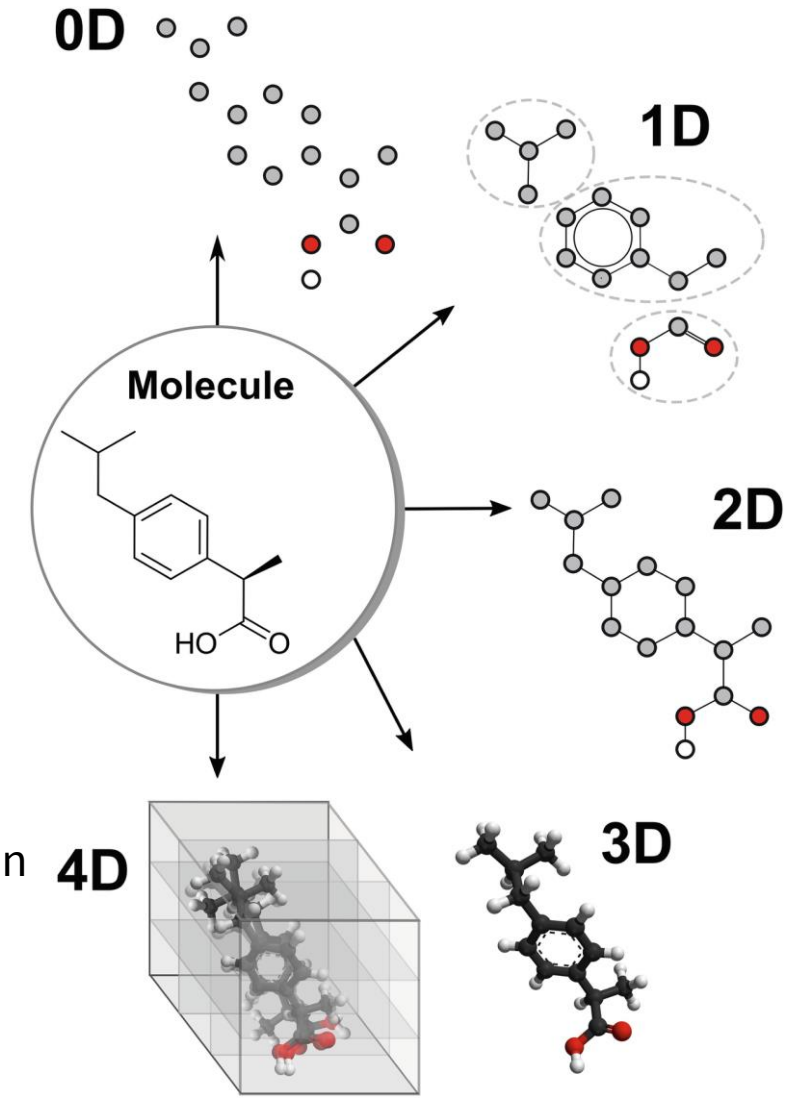
BCL::Mol2D–a robust atom environment descriptor for QSAR modeling and lead optimization

Oanh Vu

Meiler Lab

Atomic descriptors in ligand-based drug discovery

- 2D: how the atoms are connected, in terms of presence and nature of chemical bonds
- 3D: spatial geometry object in space and, in addition to the nature and connectivity of the atoms
- 4D: quantitatively identify and characterize the interactions between a molecule and a receptor's binding site



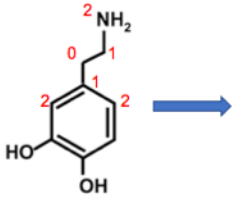
2D-fingerprint-based models perform as well as the state-of-the-art 3D structure-based models for the predictions of toxicity, solubility, partition coefficient and protein–ligand binding affinity based on only ligand information. (***Phys. Chem. Chem. Phys.***, 2020,**22**, 8373-8390)

Grisoni F., Ballabio D., Todeschini R., Consonni V. (2018). Methods in Molecular Biology, vol 1800.

Atom environment-based descriptors are shown to be most robust among 2D descriptors

fingerprint	average of all settings EF(1%)	best single setting EF(1%)
dendritic	16.2	34.7
linear	14.5	33.5
MACCS	7.3	21.6
MOLPRINT2D	22.2	35.1
pairwise	13.2	29.5
radial	13.3	33.8
torsion	15.3	34.0
triplet	15.3	34.9

Sastry., M et al. *J. Chem. Inf. Model.* **2010** *50* (5), 771-784



Layers	0	1	2
	C	=C	-N
		-C	=C
		-C	-C

Layers	0	1	2
	C	-C	-N
		-C	~C
			~C

Bender et al. J Chem Inf Comput Sci. 2004;44(5):1708-18.

BCL:: Mol2D – an improved version of Molprint2D

Characteristics	Molprint2D	BCL::Mol2D
AE Layer #	2	1
Atomic encoding	Element Bond order	Element Bond order Hybridization (Atom type)
AE Value type	Presence	Count

Generation of the master AE list

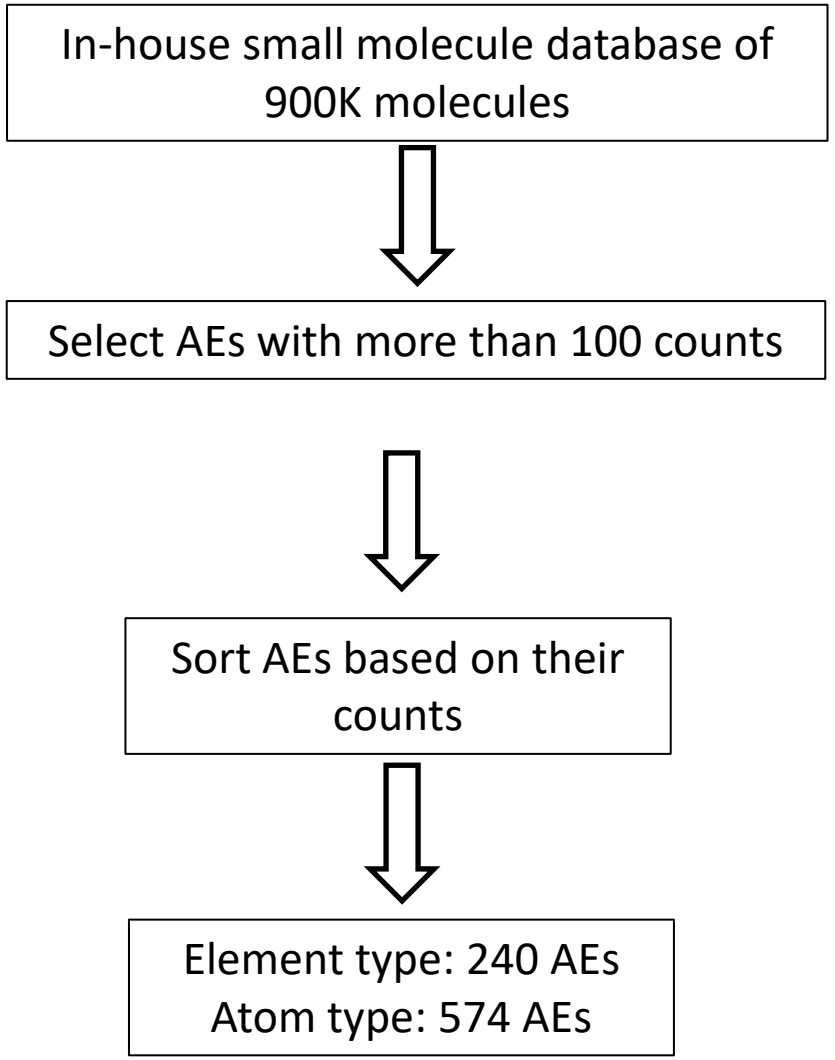
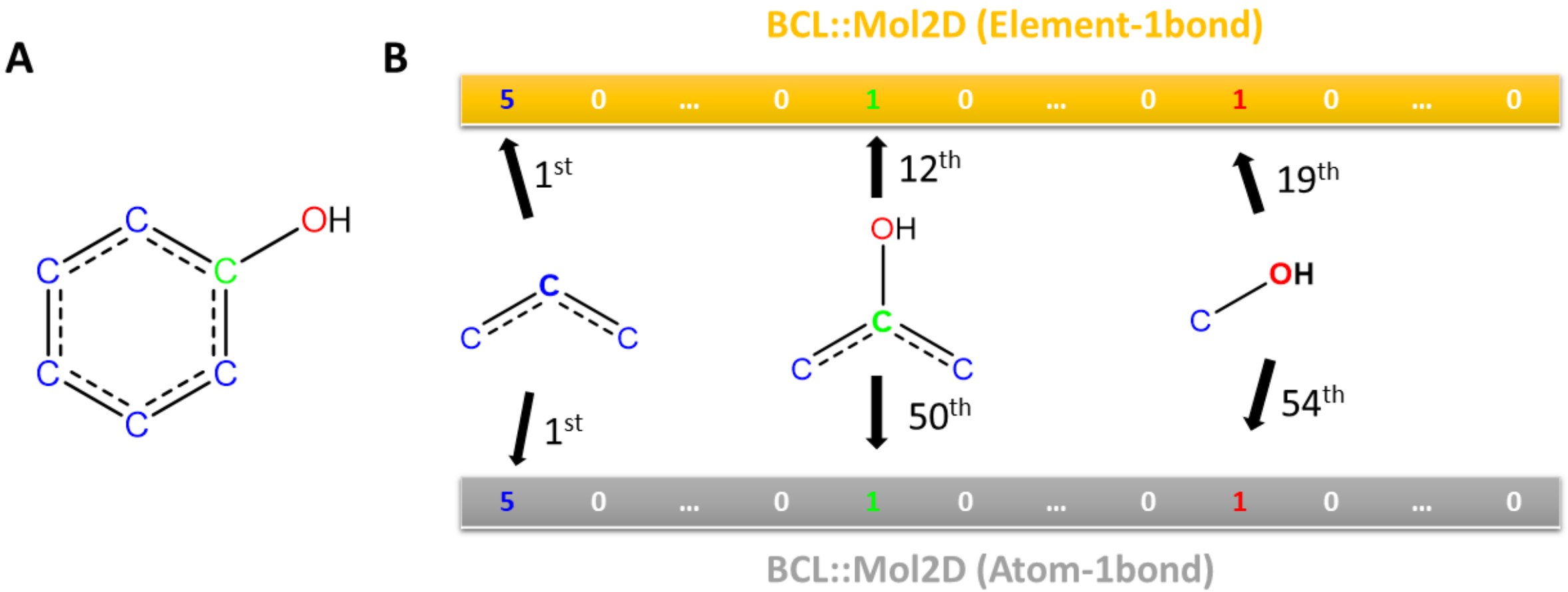
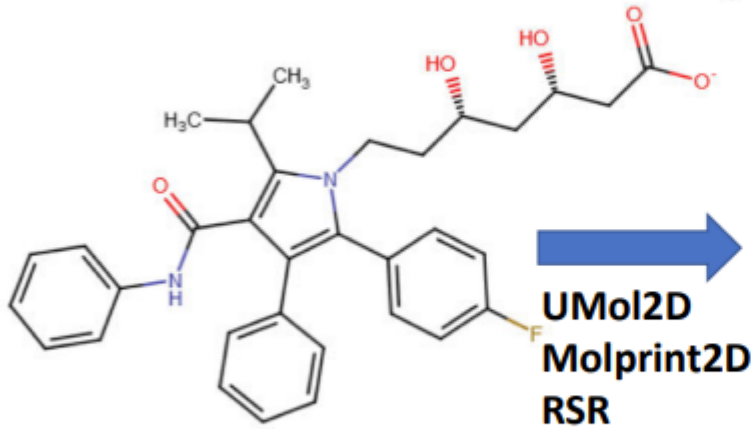


Illustration of BCL::Mol2D fingerprints of Phenol

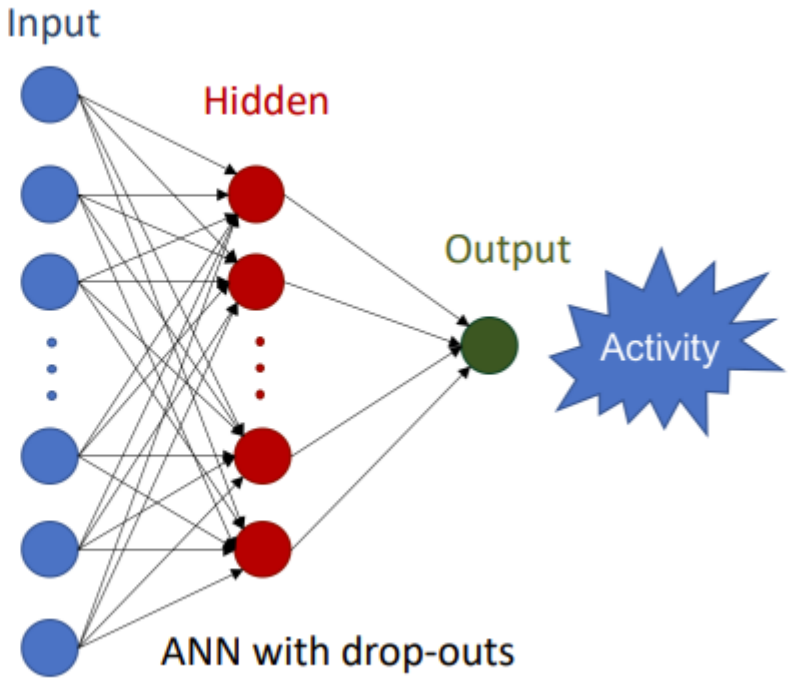


Benchmark protocol

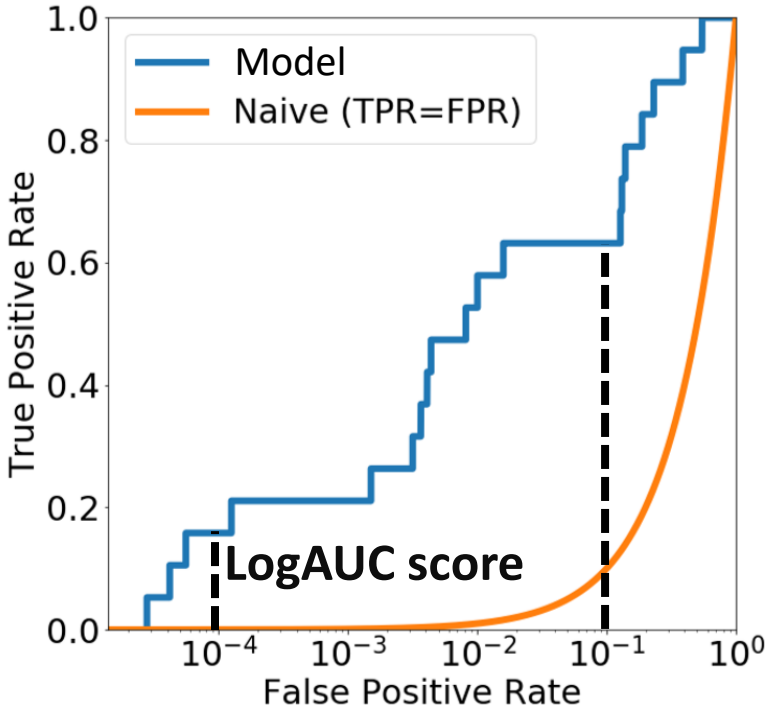


9 PubChem HTS datasets

UMol2D
Molprint2D
RSR



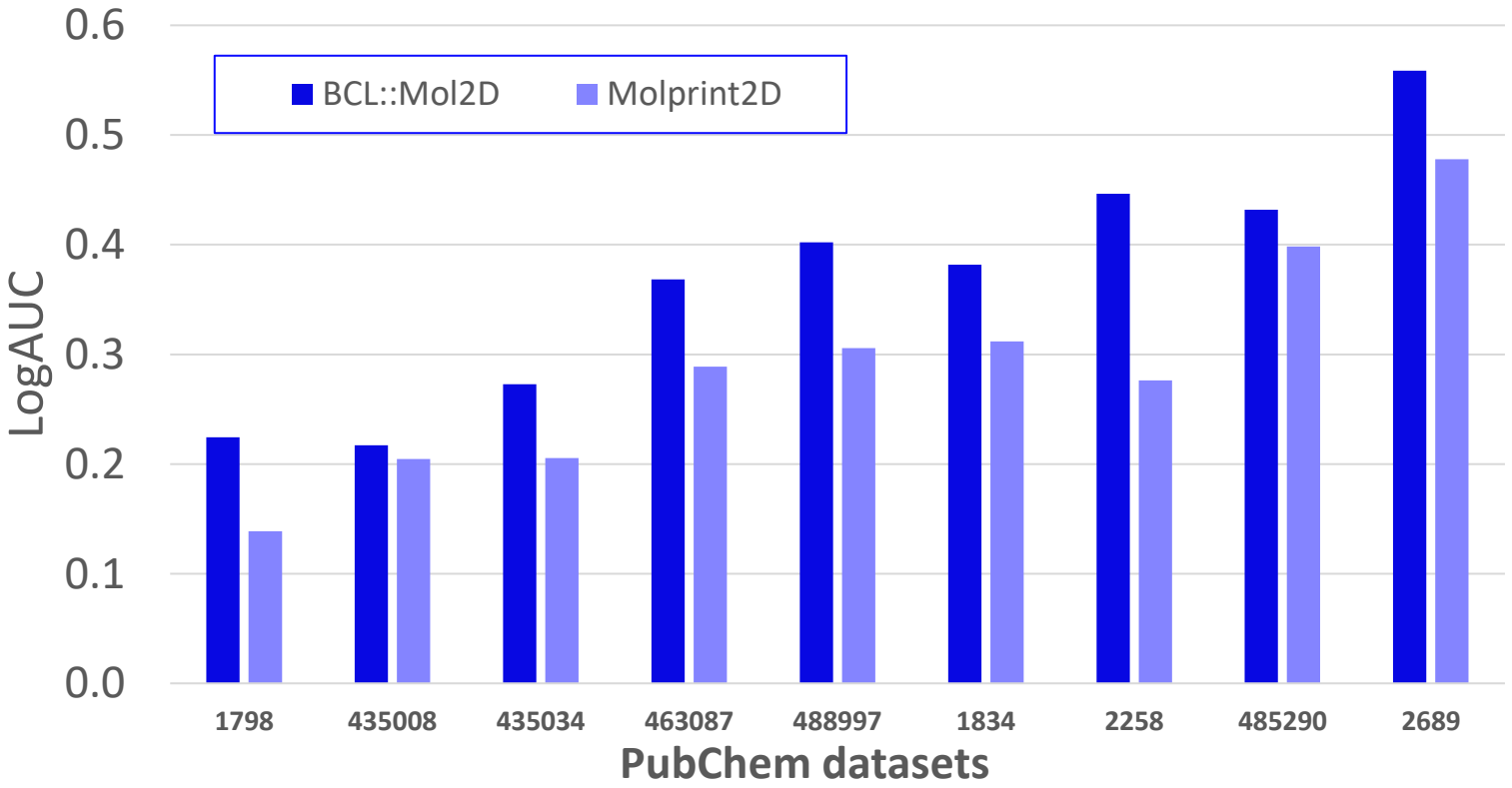
5-fold cross validation



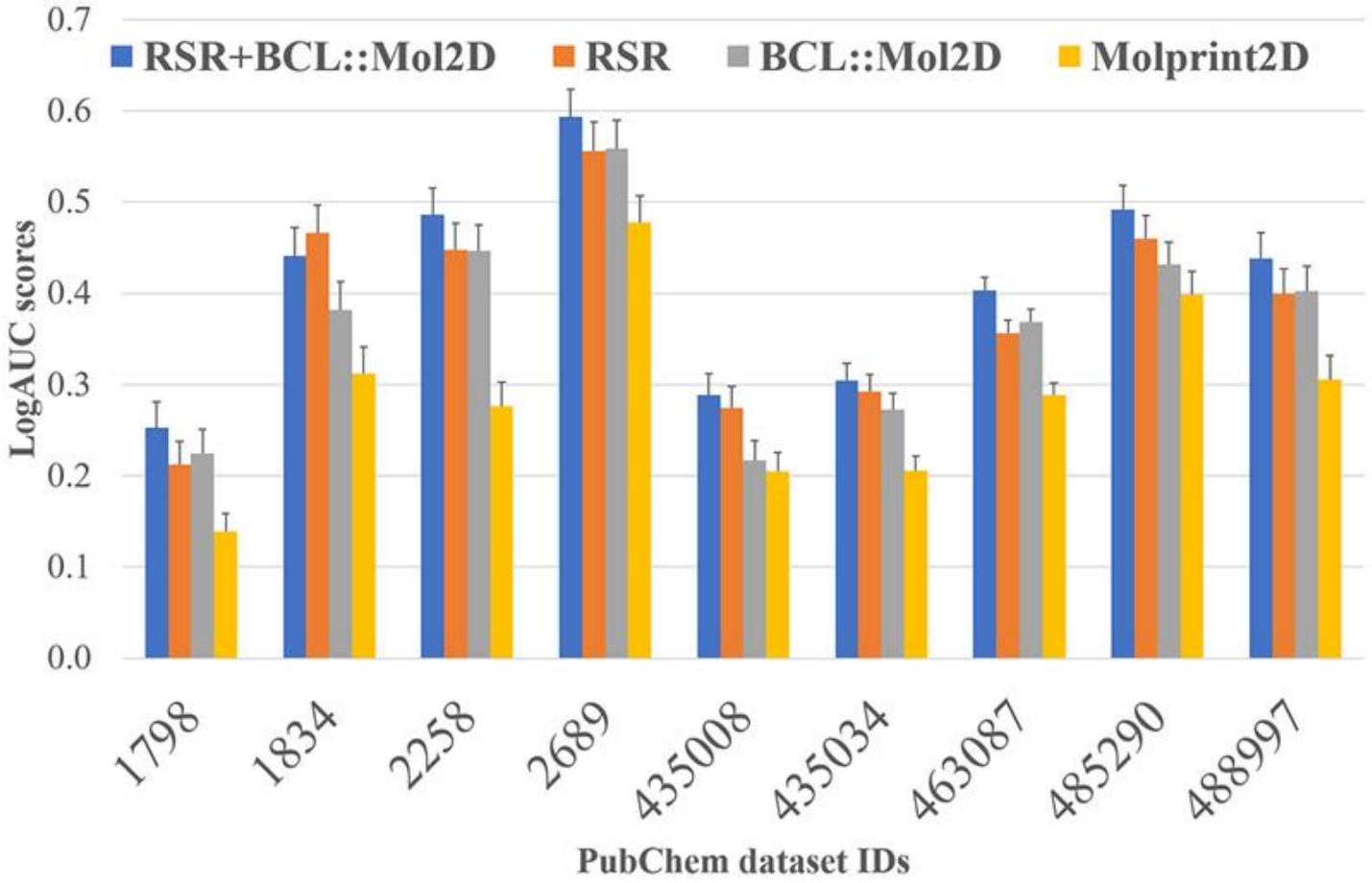
Mendenhall J, Meiler J. J Comput Aided Mol Des. 2016 Feb;30(2):177-89.

Mysinger & Shoichet. J *Chem Info Modeling* **2010,** *50* (9), 1561-1573.

BCL::Mol2D outperform Molprint2D in QSAR performance



Compare with RSR sets

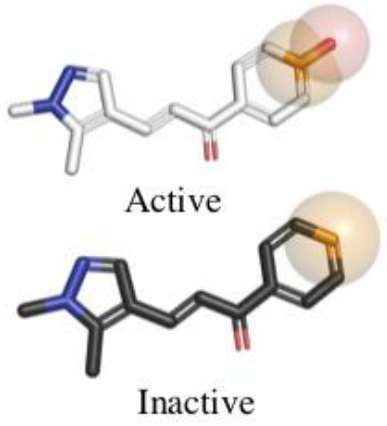


Individual AE Contributions to ANN Prediction-Sensitivity Analysis

>90% structural similarity
9 pairs

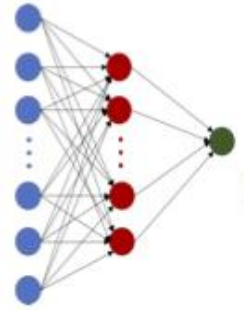
Decrement sensitivity score

$$S^{m}_{f(d_i),d_i} = \frac{f(d_i - \delta) - f(d_i)}{\delta}$$



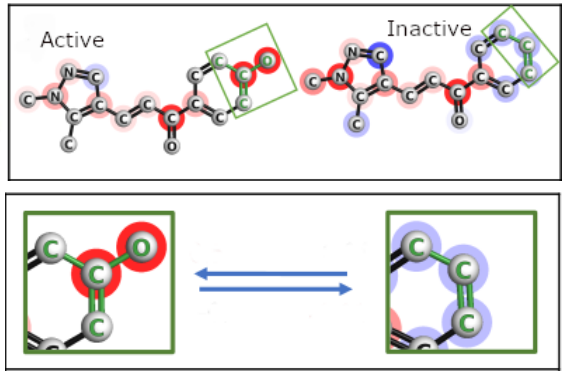
STK33 antagonists

BCL::Mol2D



QSAR-ANNs

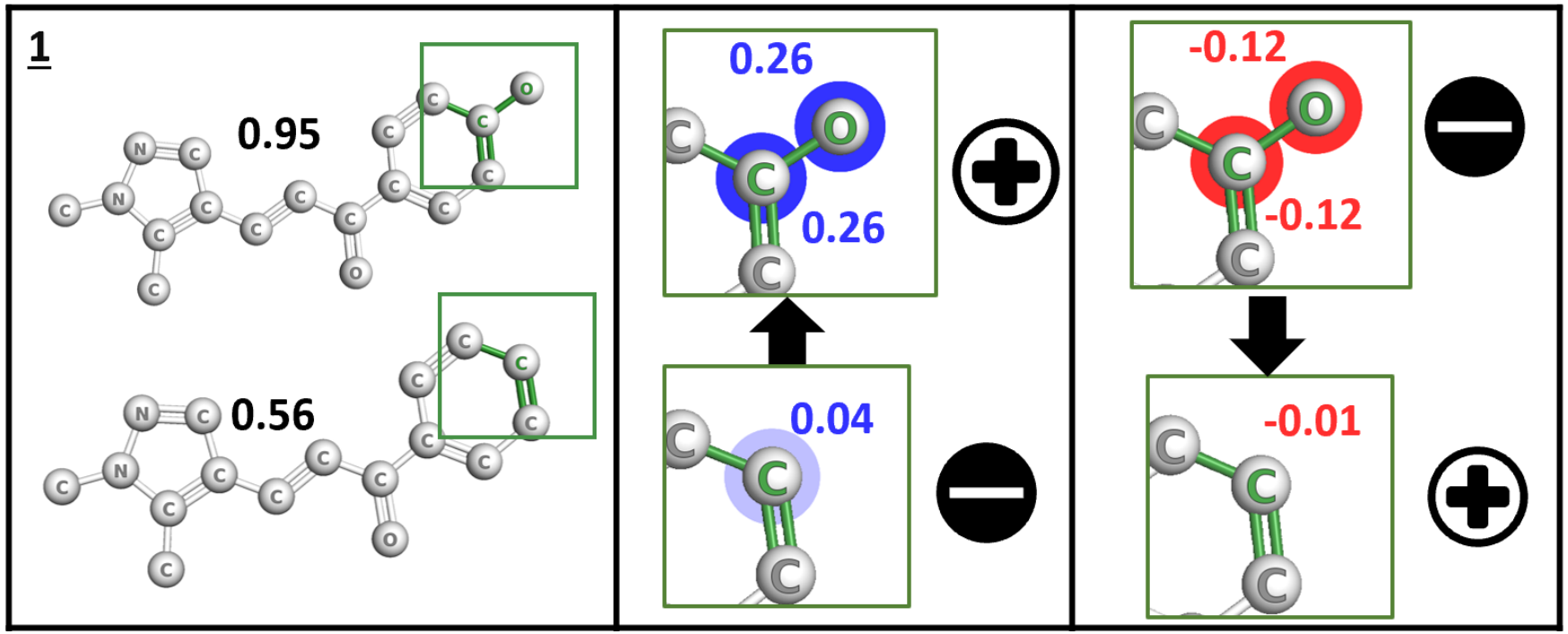
Sensitivity analysis



Favorable

Not favorable

Mapping partial contributions of AEs to active and inactive compounds



Decrement sensitivity score

$$S^{m}_{f(d_i),d_i} = \frac{f(d_i - \delta) - f(d_i)}{\delta}$$

Increment sensitivity score

$$S^{m}_{f(d_i),d_i} = \frac{f(d_i + \delta) - f(d_i)}{\delta}$$

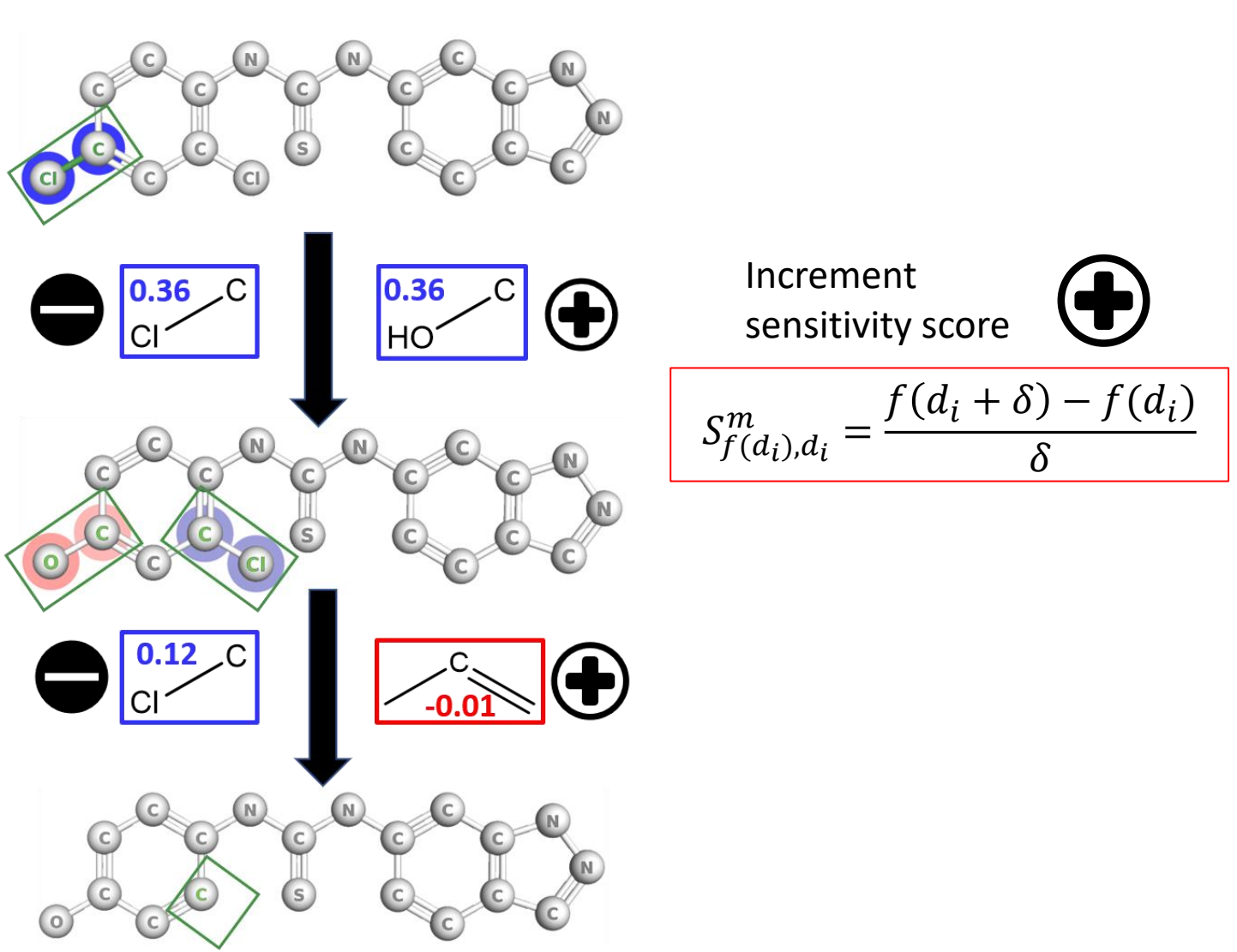
BCL::Mol2D sensitivity score can guide lead optimization

Decrement sensitivity score

$$S^{m}_{f(d_i),d_i} = \frac{f(d_i - \delta) - f(d_i)}{\delta}$$

Increment sensitivity score

$$S^{m}_{f(d_i),d_i} = \frac{f(d_i + \delta) - f(d_i)}{\delta}$$



Conclusion

- Implement and incorporate the BCL::Mol2D descriptor in to QSAR pipeline in BCL::ChemInfo
- ANN models trained on BCL:Mol2D showed improve prediction over that trained on Molprint2D
- BCL::Mol2D sensitivity analysis can guide lead optimization

Acknowledgement

The Meiler lab:
Jens Meiler
Jeffrey Mendenhall
Benjamin Brown
Everyone else