

Exploratory and Explanatory Visualization of Molecules and Chemical Models

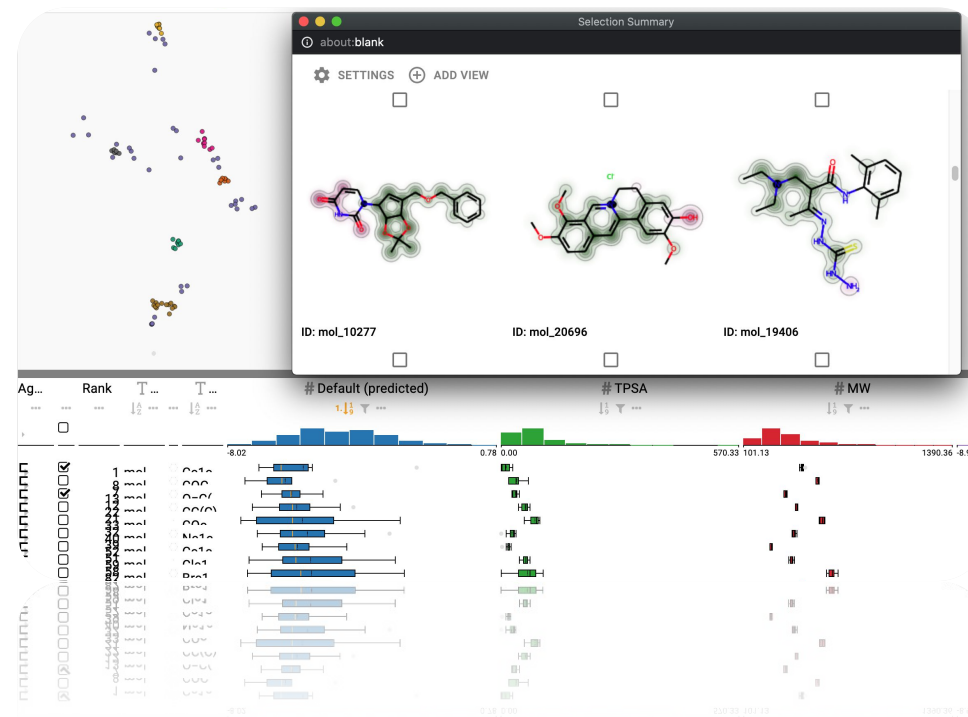
with **CIME**,
the ChemInformatics Model Explorer



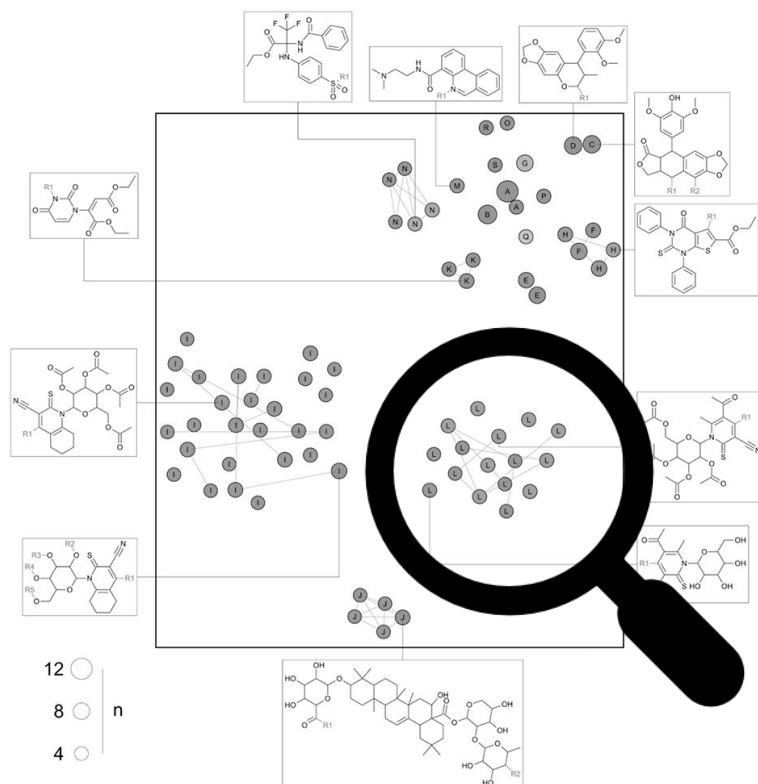
Henry Heberle
Data Science (Visualization)
Germany



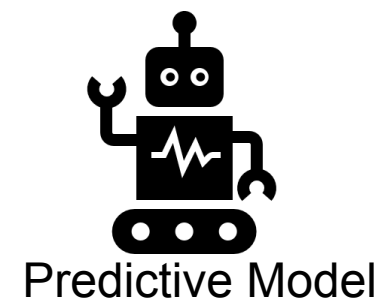
Christina Humer
Data Science (Visualization)
Austria



Machine Learning models can predict molecules' properties



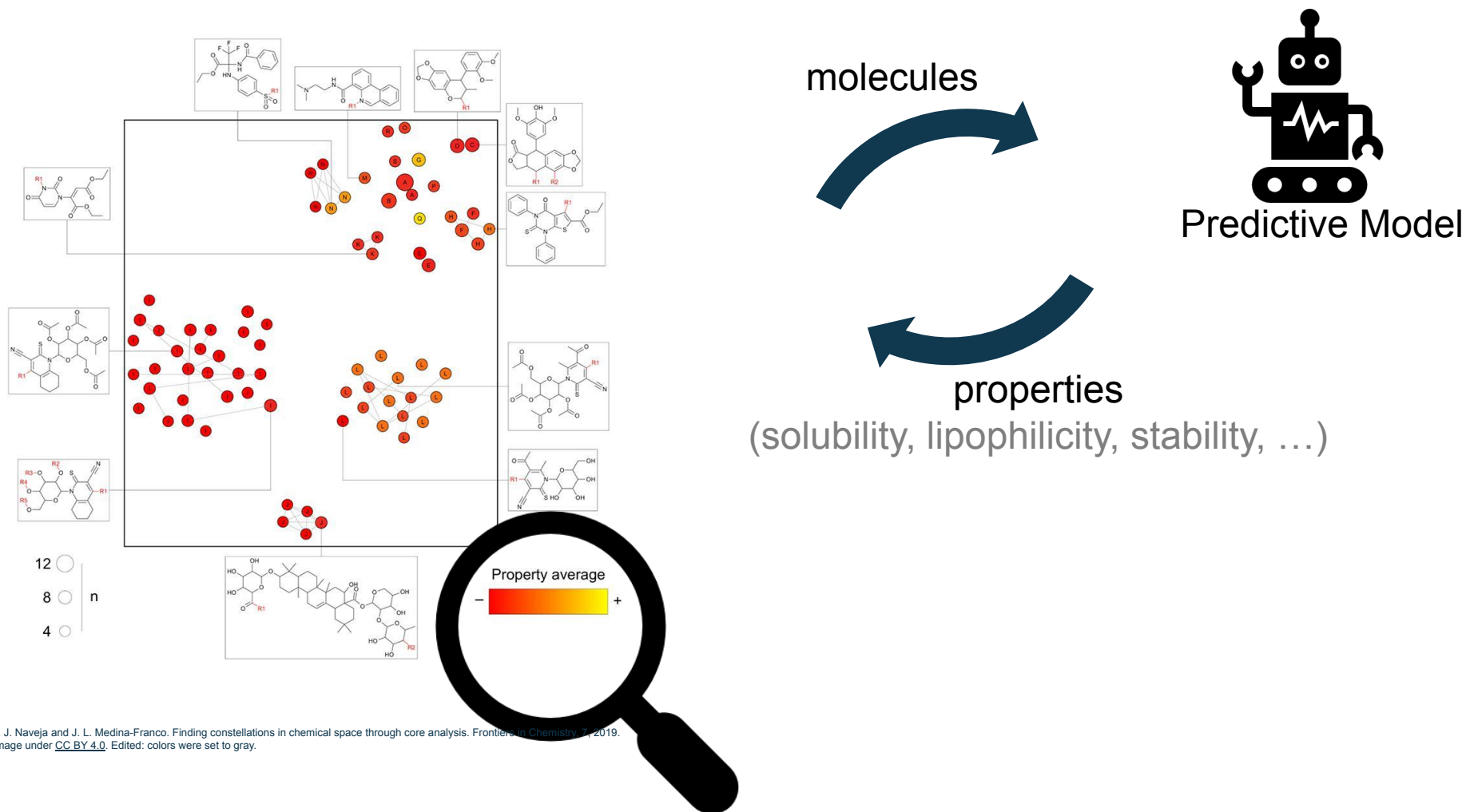
molecules



Which ones are soluble in water?

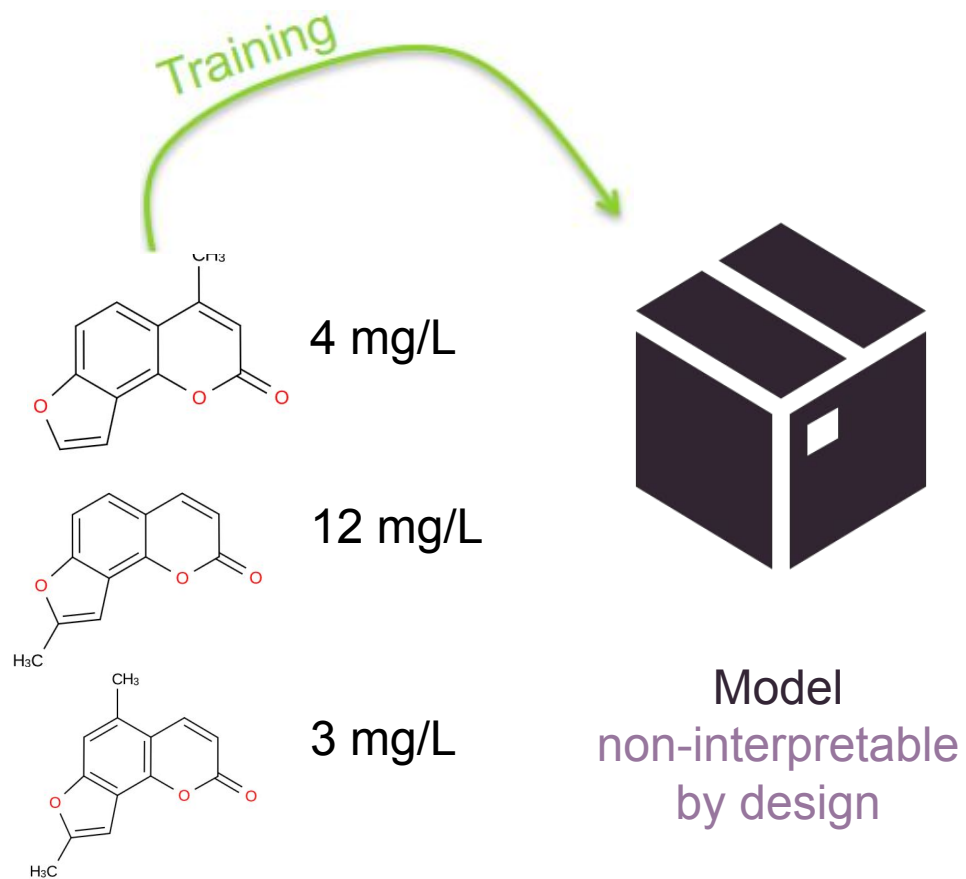
J. J. Naveja and J. L. Medina-Franco. Finding constellations in chemical space through core analysis. *Frontiers in Chemistry*, 7, 2019.
Image under [CC BY 4.0](#). Edited: colors were set to gray.

Machine Learning models can predict molecules' properties

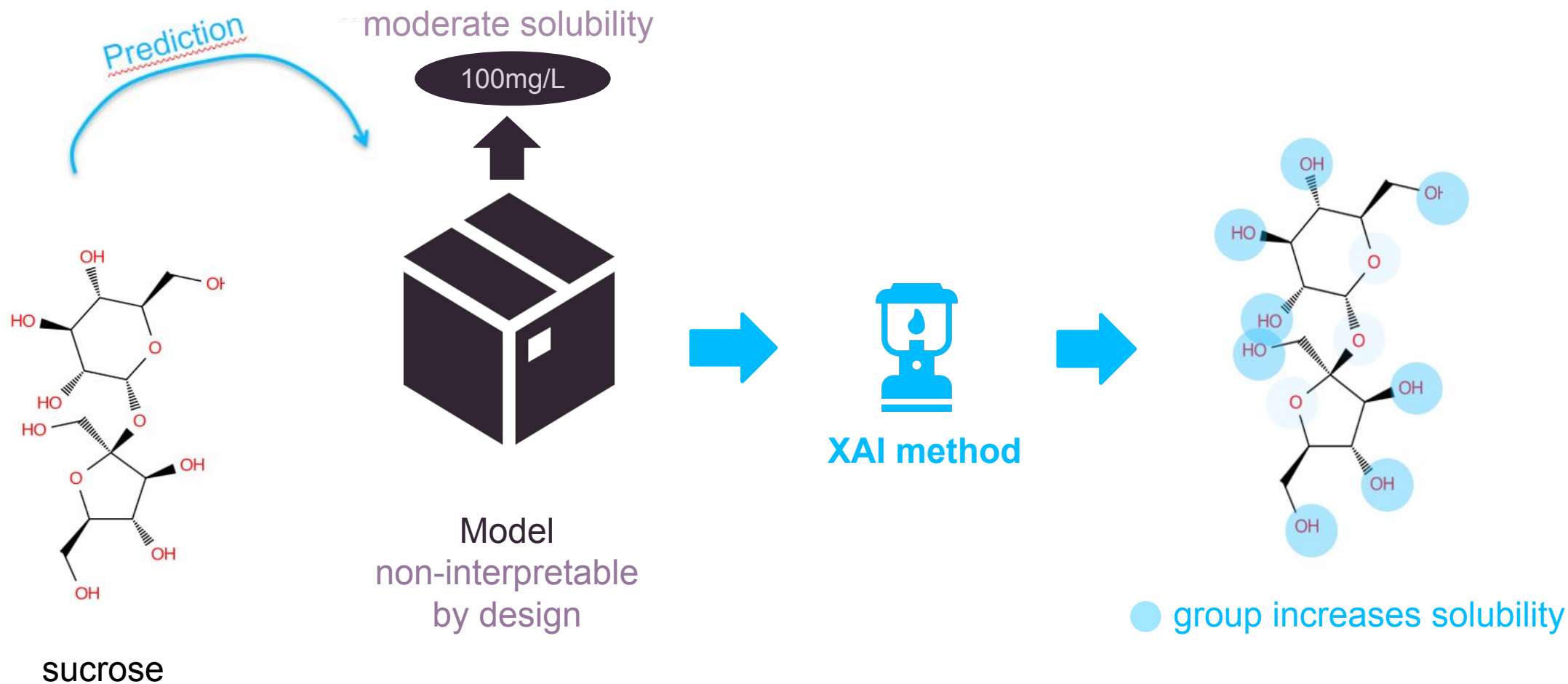


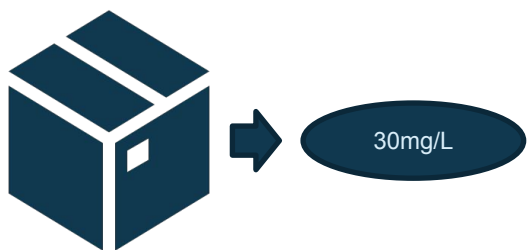
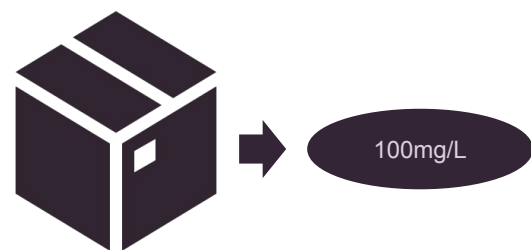
J. J. Naveja and J. L. Medina-Franco. Finding constellations in chemical space through core analysis. *Frontiers in Chemistry*, 7, 2019. Image under [CC BY 4.0](#). Edited: colors were set to gray.

Explainable AI (XAI) techniques highlight important regions

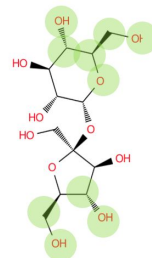
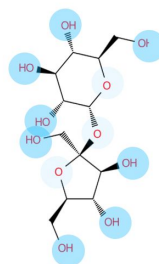


Explainable AI (XAI) techniques highlight important regions

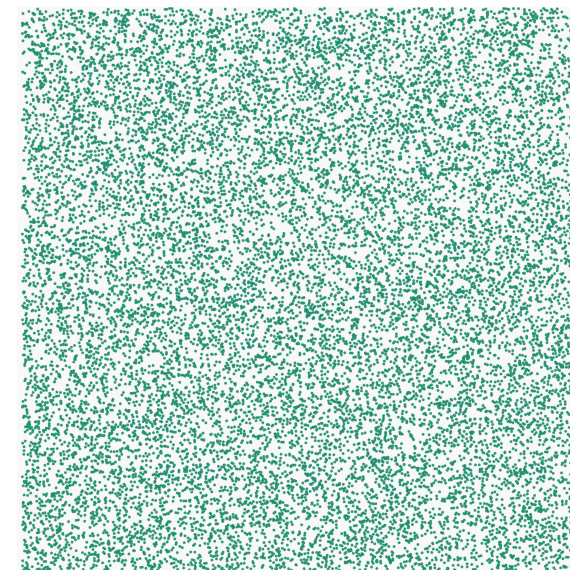




Property predictions



Multiple explanations
from different methods
or properties



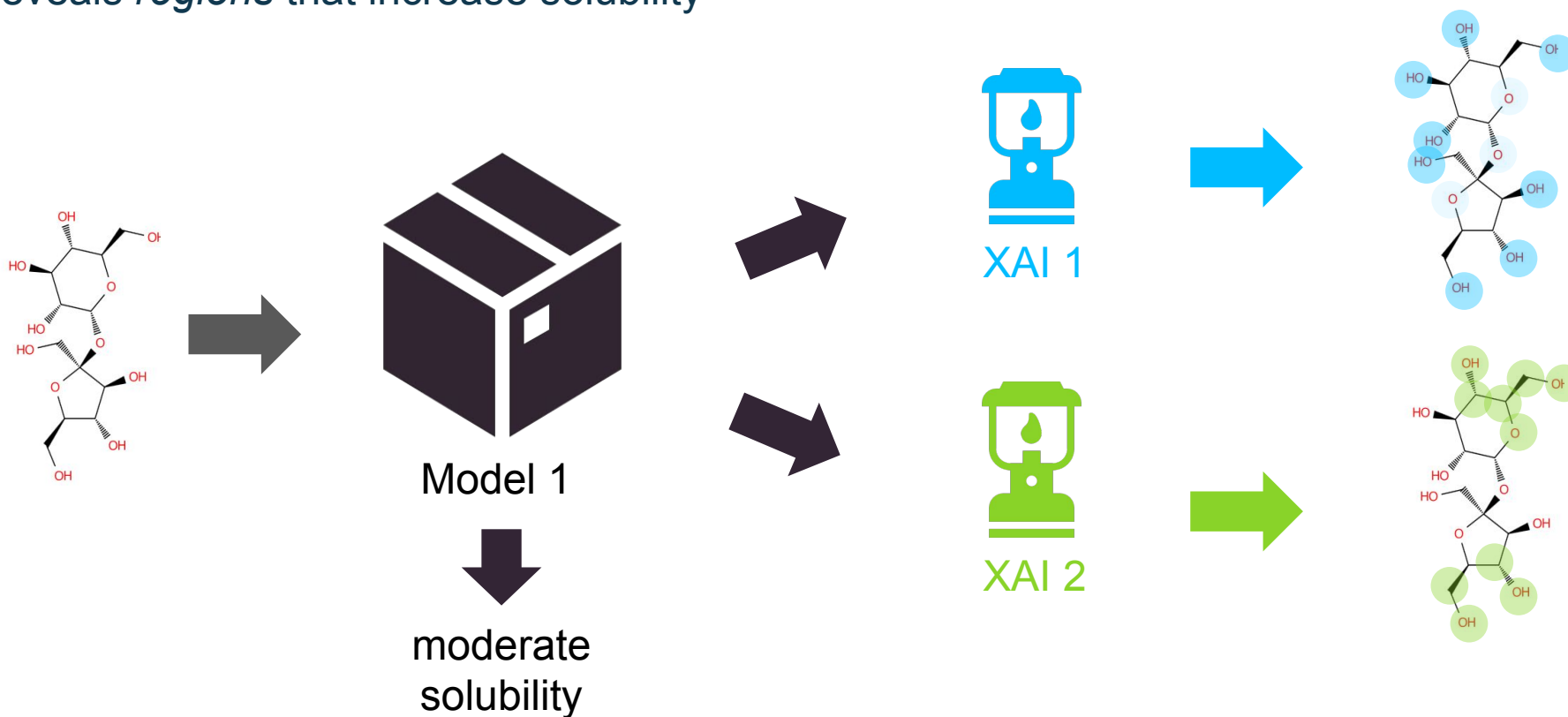
Big chemical space

Exploratory Visualization

Example of task: comparing two XAI methods

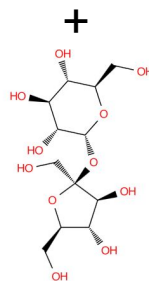
Suppose we want a model to predict solubility that:

- is *accurate*
- reveals *regions* that increase solubility



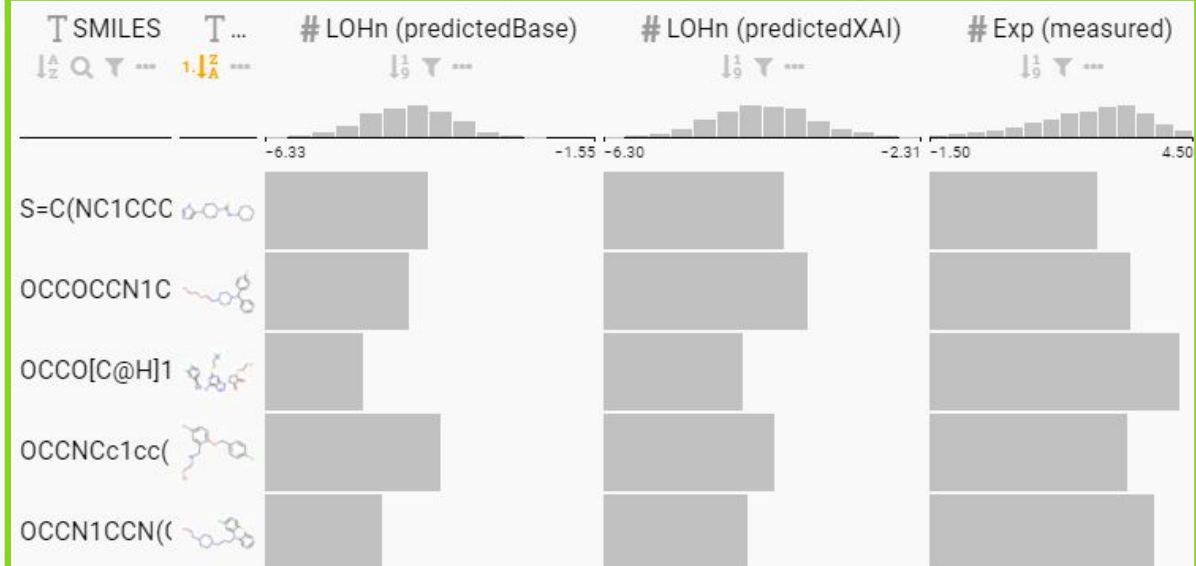
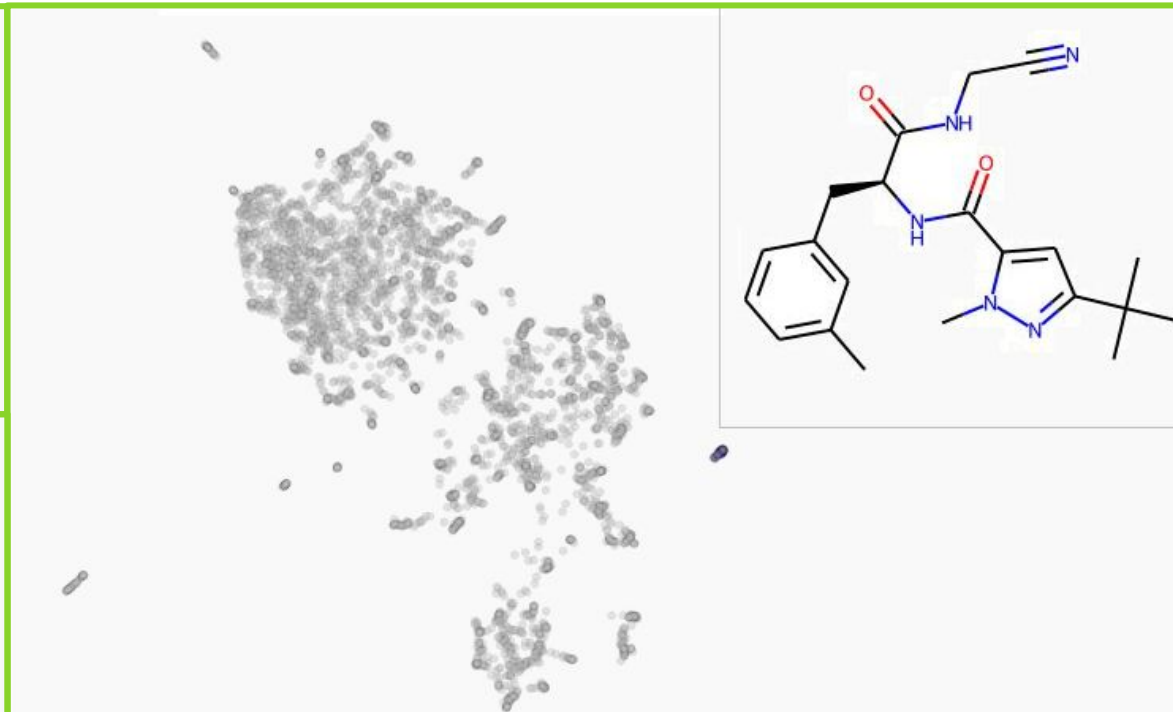
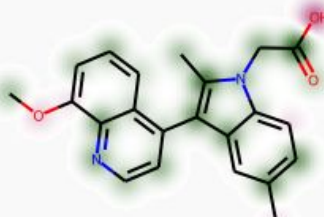
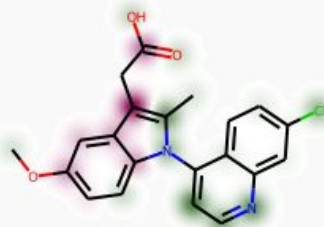
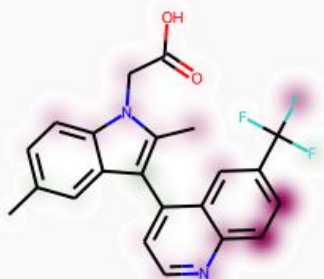
CIME

(X)AI



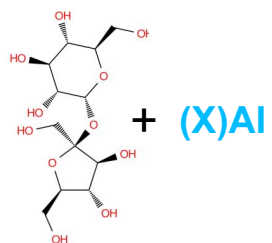
OPEN FILE

Drop a file here



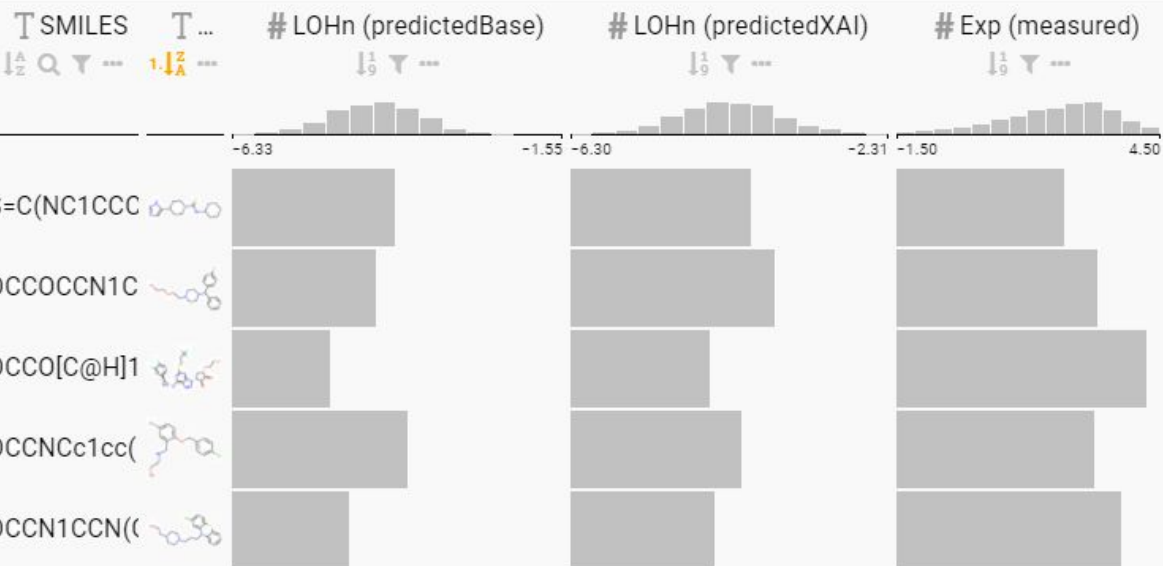
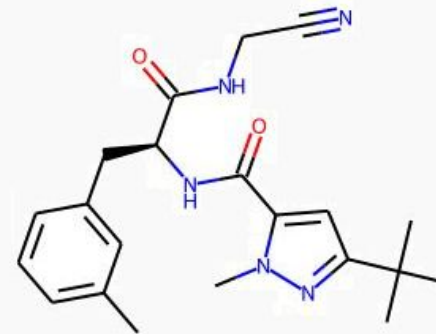
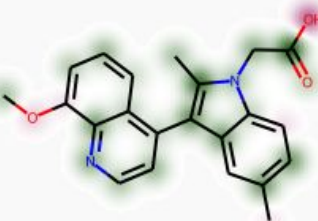
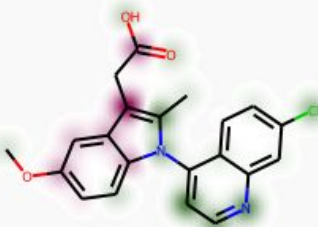
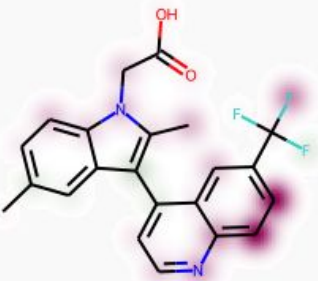
CIME

SDF



OPEN FILE

Drop a file here



Example Dataset (SDF)

Atoms

Atom Bonds

Properties

xAI

```
RDKit      2D
26 28 0 0 0 0 0 0 0 0 0999 V2000
-3.8971 -10.3573 0.0000 C ...
-3.8971 -8.8573 0.0000 C ...
-2.5981 -8.1073 0.0000 C ...
0.0000 -3.6073 0.0000 N ...
1.2135 -2.7256 0.0000 N ...
```

```
...
1 2 1 0
2 3 1 0
3 4 1 0
3 5 1 0
3 6 1 0
```

M END

```
> <fingerprint_0>
2.4448594558634795e-05
...
> <fingerprint_127>
0.3827087283134461
```

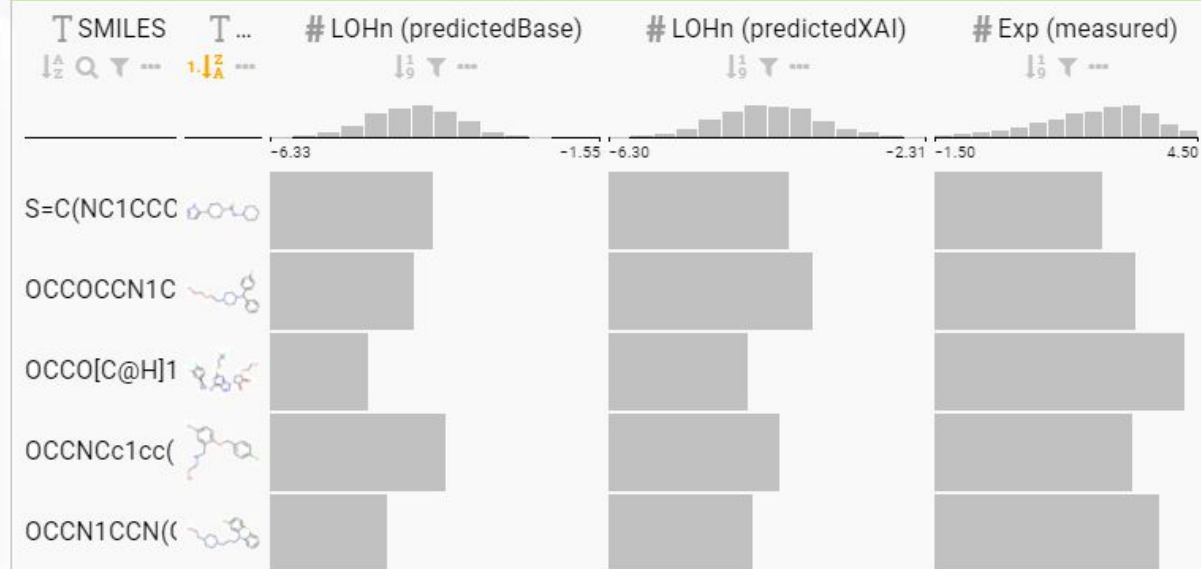
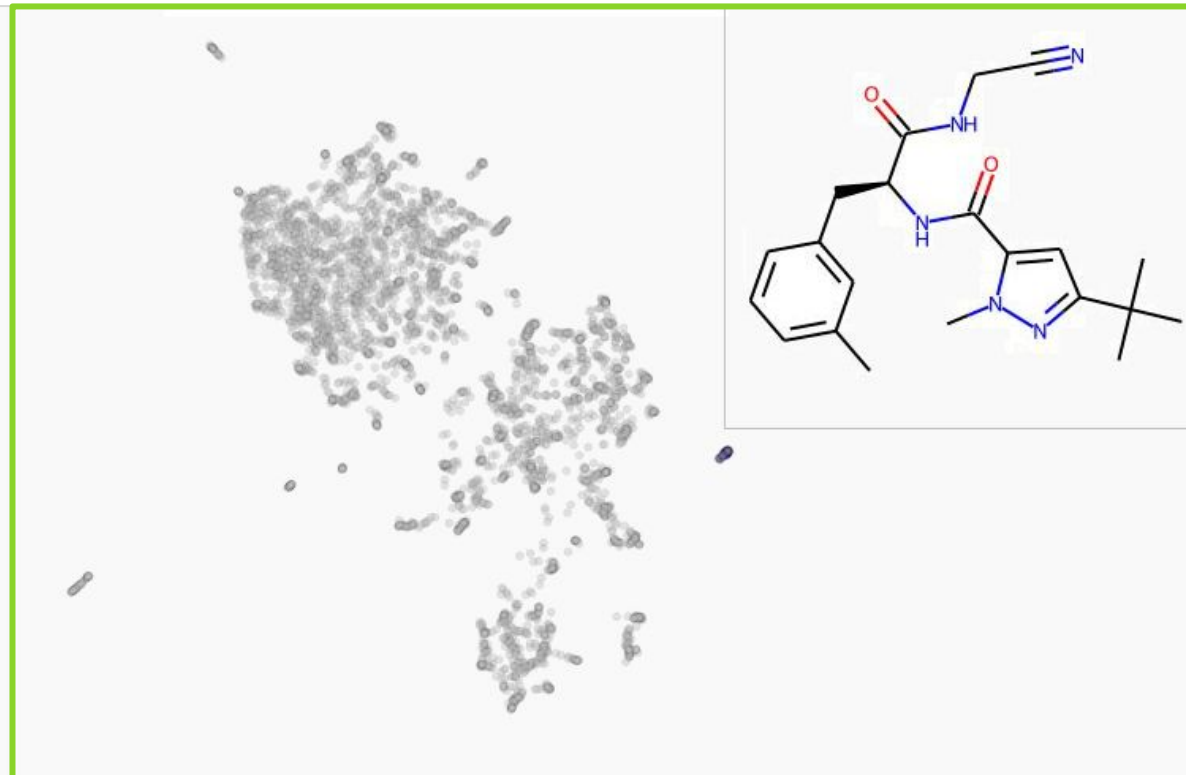
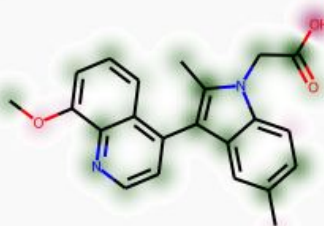
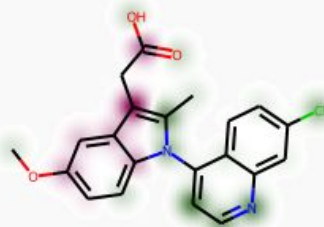
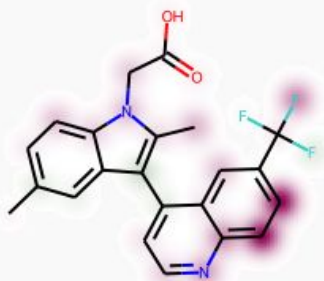
```
> <predicted_LOD>
3.2429239749908447
...
```

```
> <measured_LOD>
2.98
...
```

```
> <atom.dprop.rep_1>
-0.746948 -0.746948 -0.1121631 -0.1121631 -0.4463508 ...
> <atom.dprop.rep_2>
1.1985533 1.1985533 1.1985533 1.4049025 -0.5853193 ...
...
```

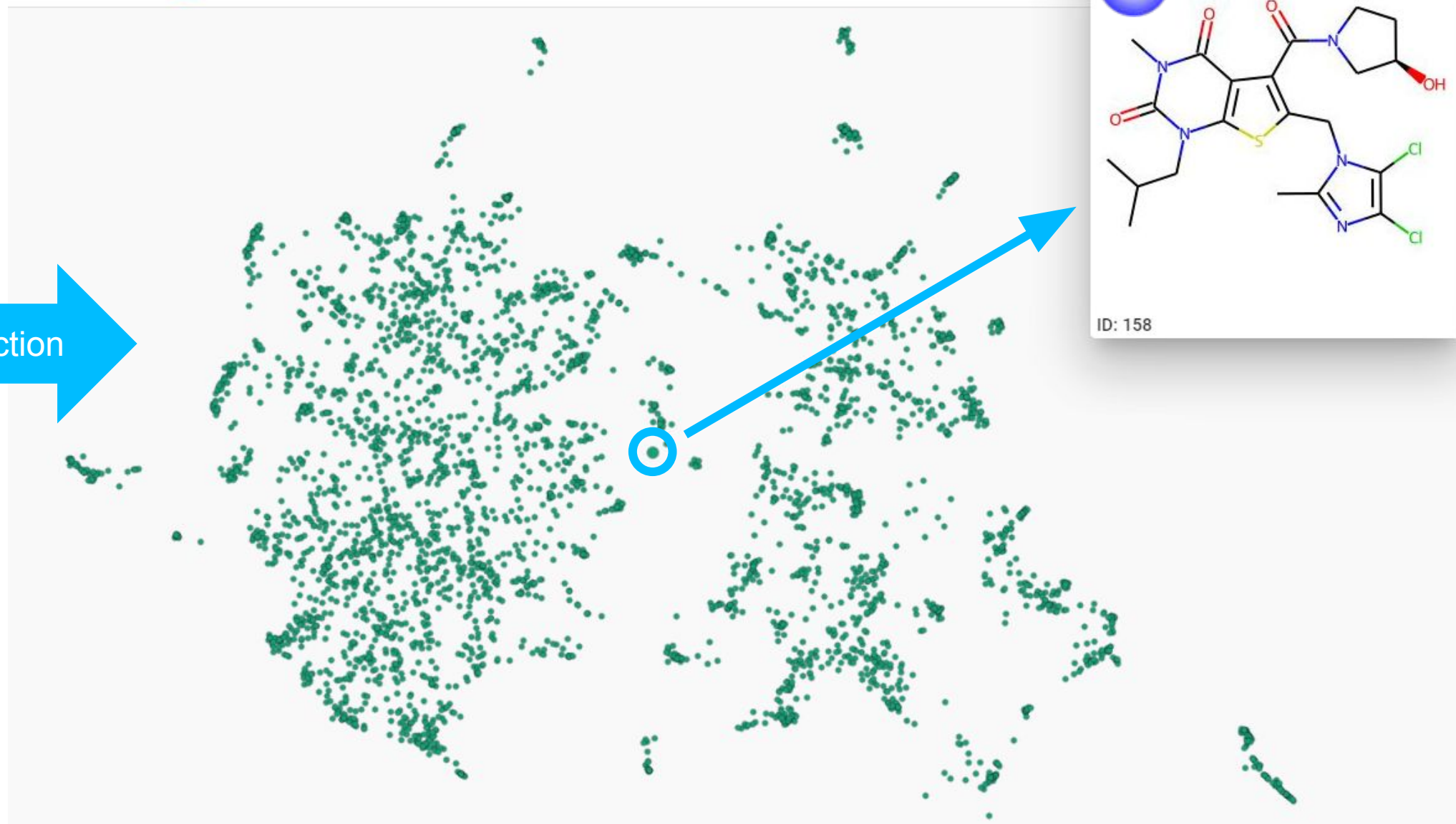
OPEN FILE

Drop a file here



Scatter View

Projection



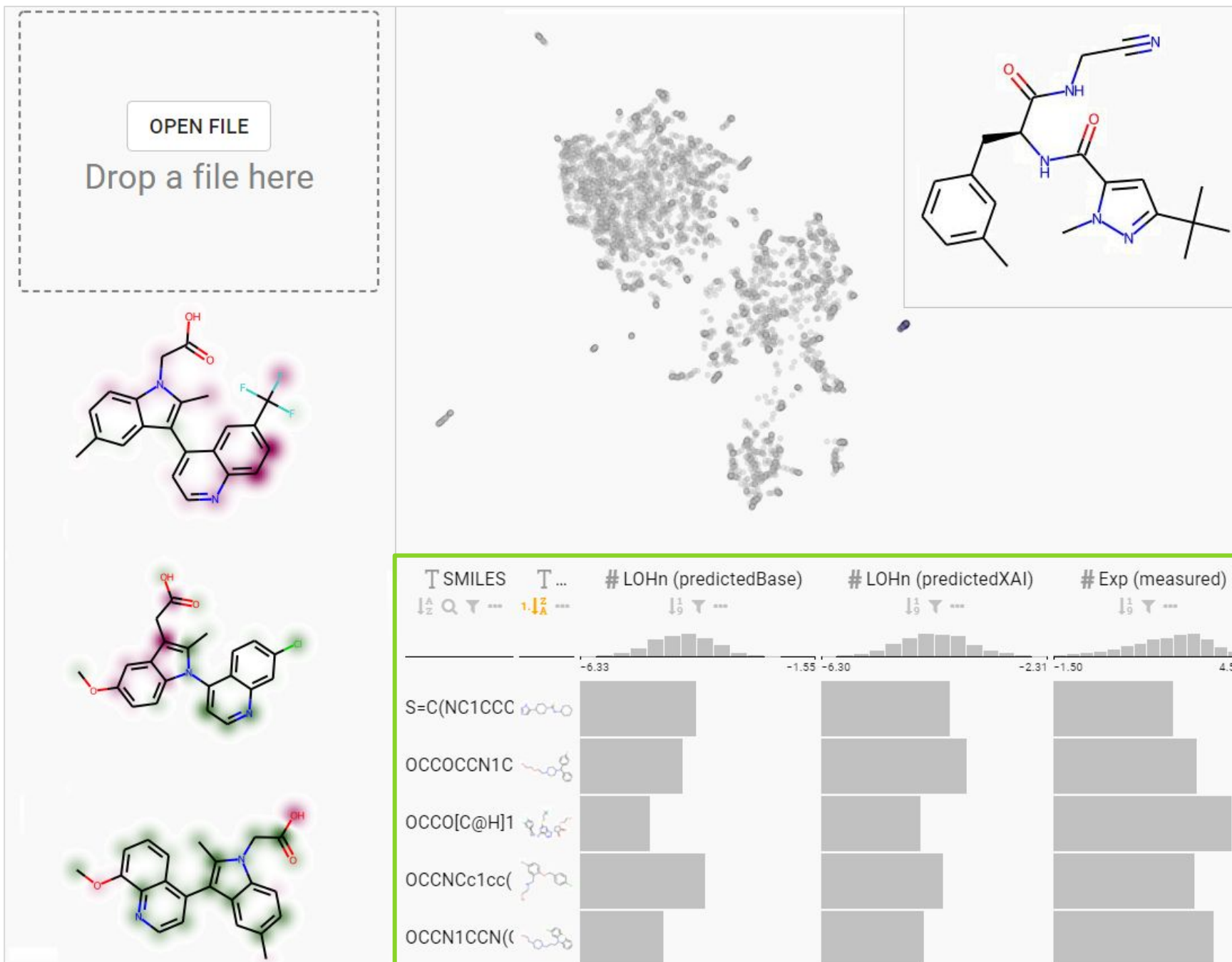
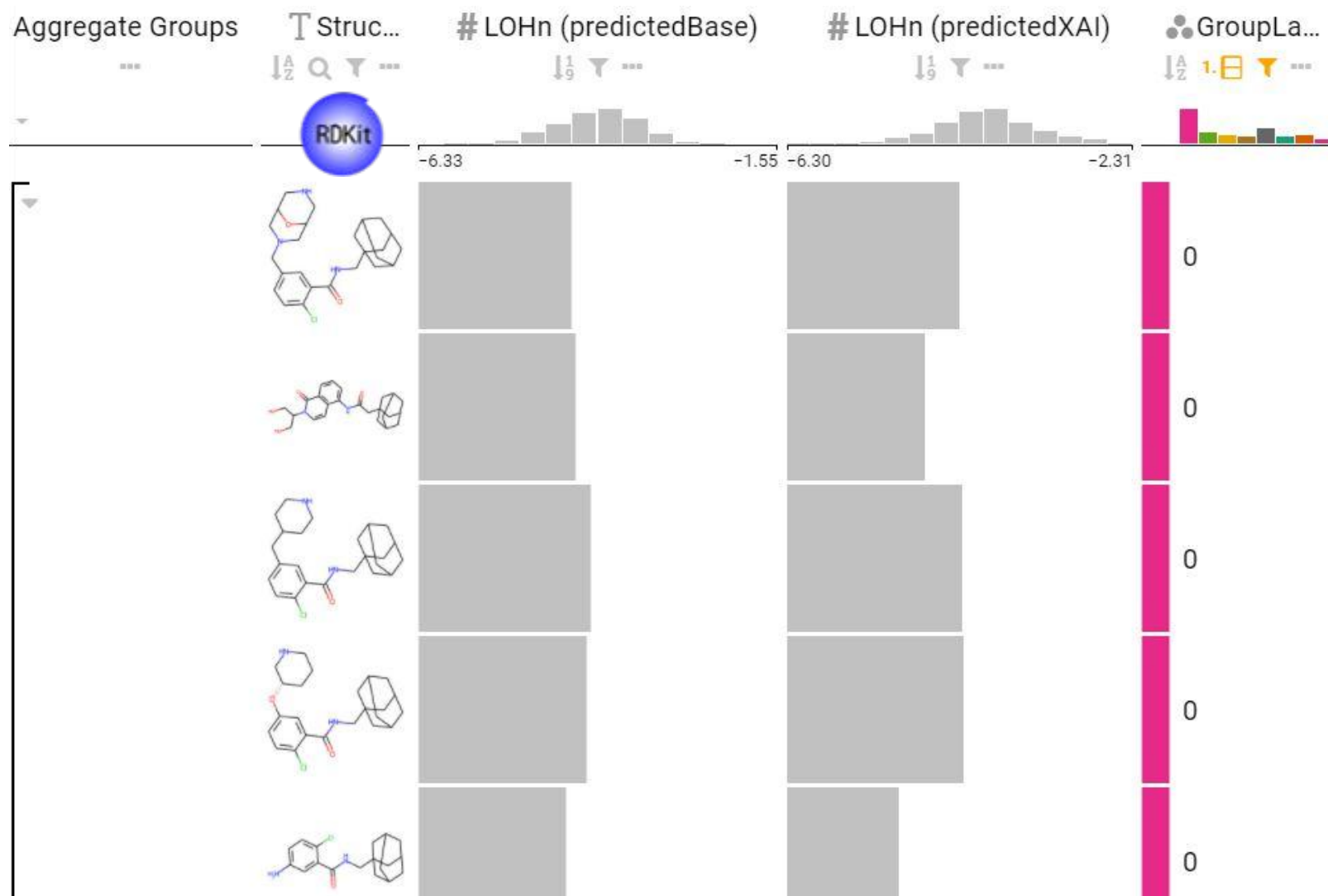


Table View



Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. **LineUp**: Visual Analysis of Multi-Attribute Rankings
 IEEE Transactions on Visualization and Computer Graphics (InfoVis '13), 19(12), pp. 2277–2286, [doi:10.1109/TVCG.2013.173](https://doi.org/10.1109/TVCG.2013.173), 2013.

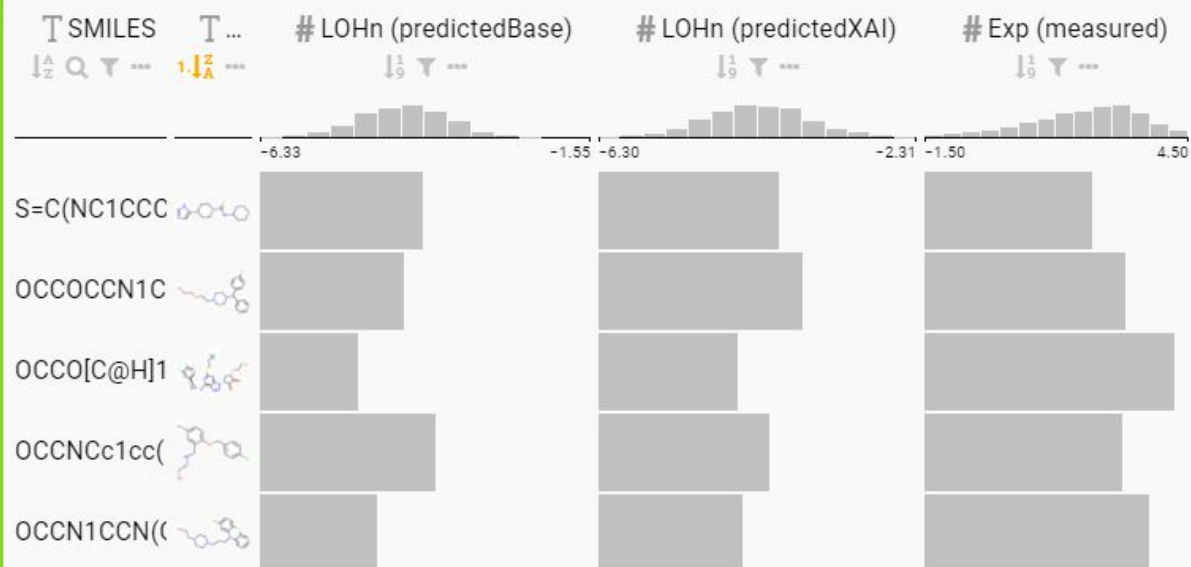
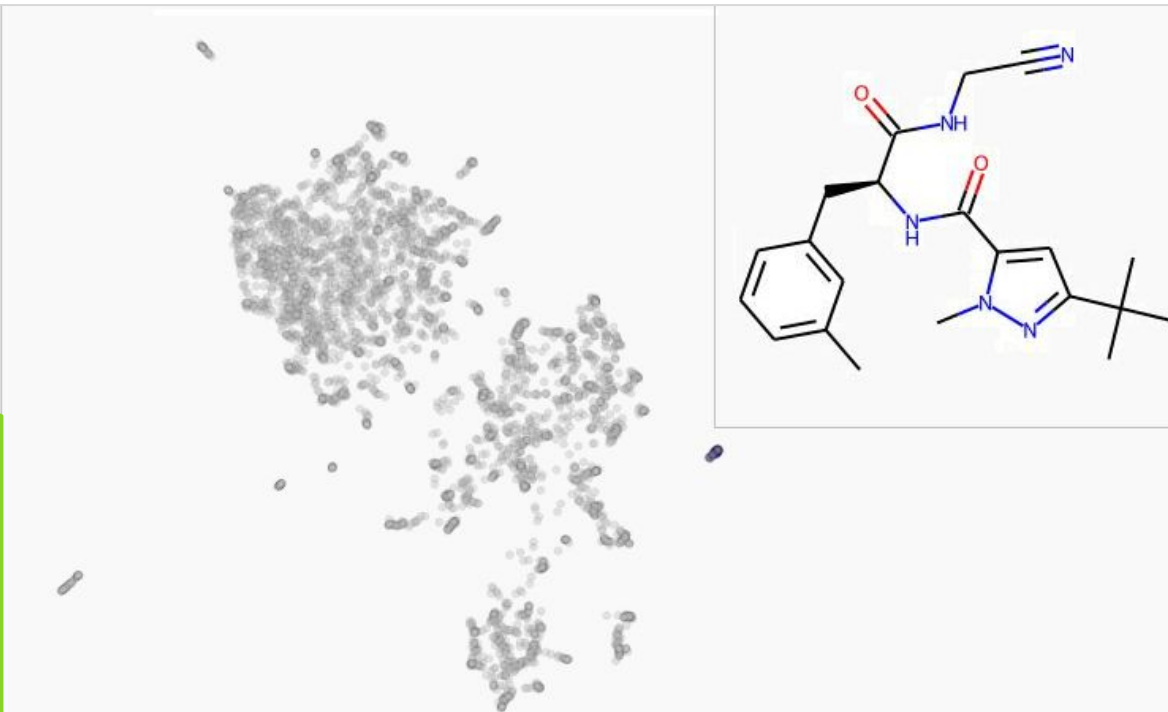
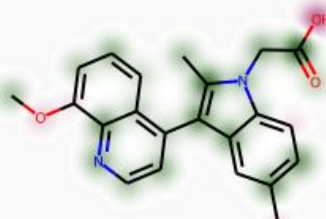
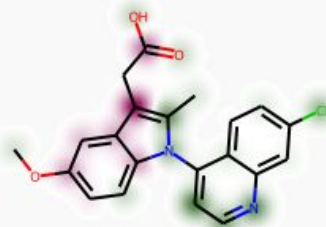
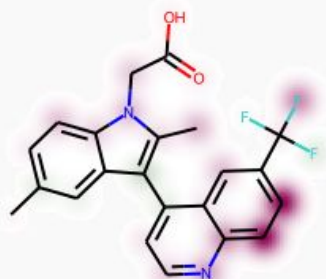
Table View



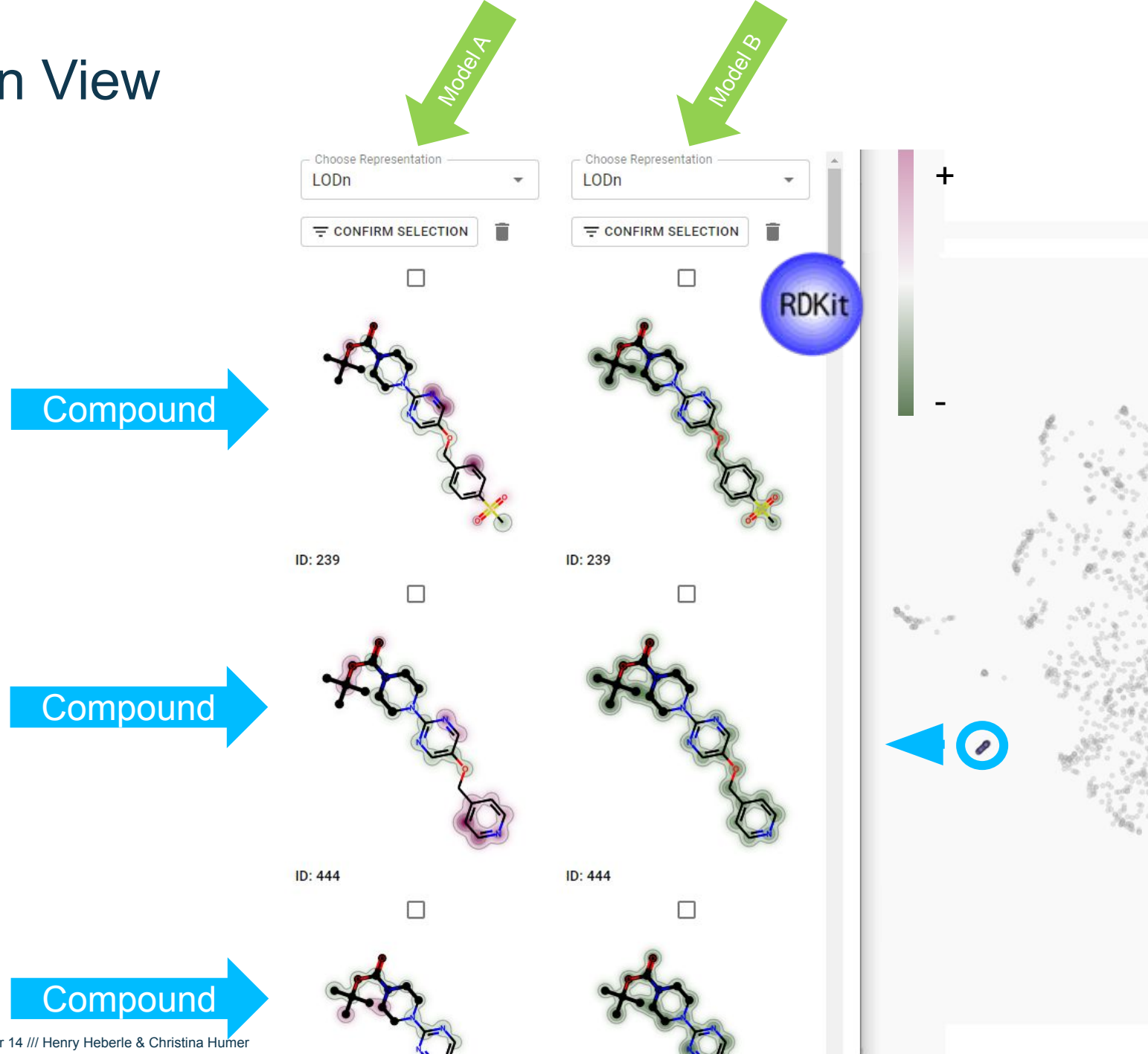
Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. **LineUp**: Visual Analysis of Multi-Attribute Rankings
 IEEE Transactions on Visualization and Computer Graphics (InfoVis '13), 19(12), pp. 2277–2286, [doi:10.1109/TVCG.2013.173](https://doi.org/10.1109/TVCG.2013.173), 2013.

OPEN FILE

Drop a file here



Contribution View



Use Case: Comparing attributions from 2 models for lipophilicity

- Lipophilicity dataset from MoleculeNet^[1]
 - 4200 molecules
- 2 models^[2] trained to predict logD



B: base model



X: designed to be more interpretable [is it?]

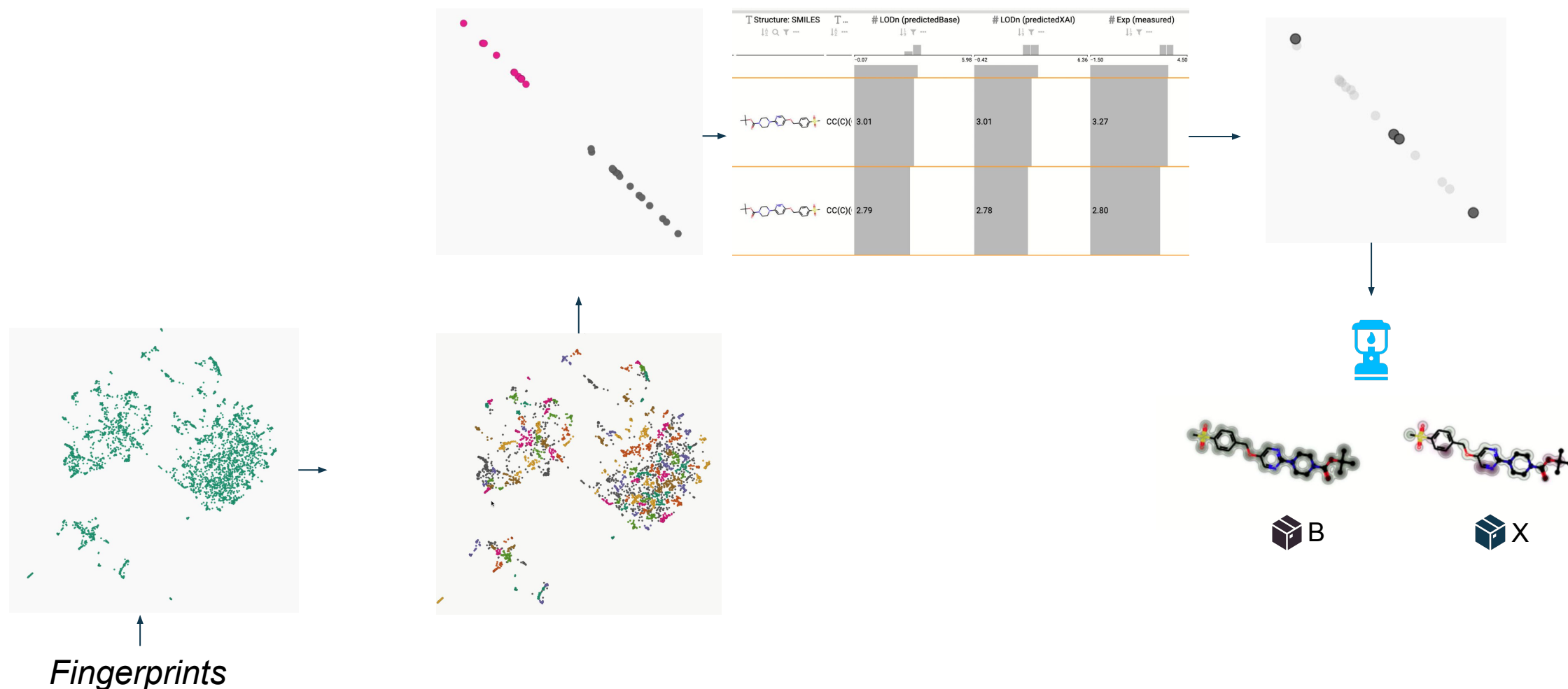


One XAI method

[1] Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., ... & Pande, V. (2018). MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, 9(2), 513-530.

[2] Henderson, R., Clevert, D. A., & Montanari, F. (2021). Improving Molecular Graph Neural Network Explainability with Orthonormalization and Induced Sparsity. *arXiv preprint arXiv:2105.04854*.

Use Case: Comparing attributions from 2 models for lipophilicity



Demo

Summary

- Model-agnostic Web Application
- Exploratory Visualization: Overview + Details
- Not only for (X)AI: any dataset with molecular or atom-level features
- Helping to:
 - Explore chemical space
 - Improve model's performance
 - Increase model's interpretability
 - Increase Trust in AI / Communicate
 - To experts
 - To Regulatory Agencies

Acknowledgments



Henry Heberle

henry.heberle@bayer.com

Julian Heinrich

Floriane Montanari

Ryan Henderson

Thomas Wolf

Florian Huber



Christina Humer

christina.humer@jku.at

Marc Streit



Greg Landrum

Demo Website

<https://jku-vds-lab.at/cime-demo>

Github

<https://github.com/jku-vds-lab/projection-space-explorer/tree/cimeV0.1.17c>

Article in preparation