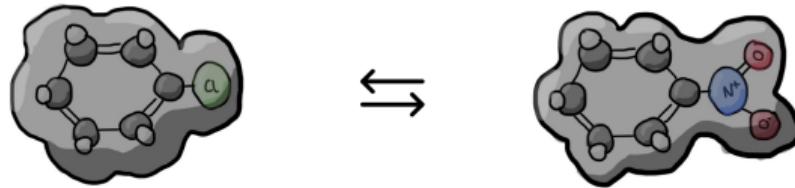
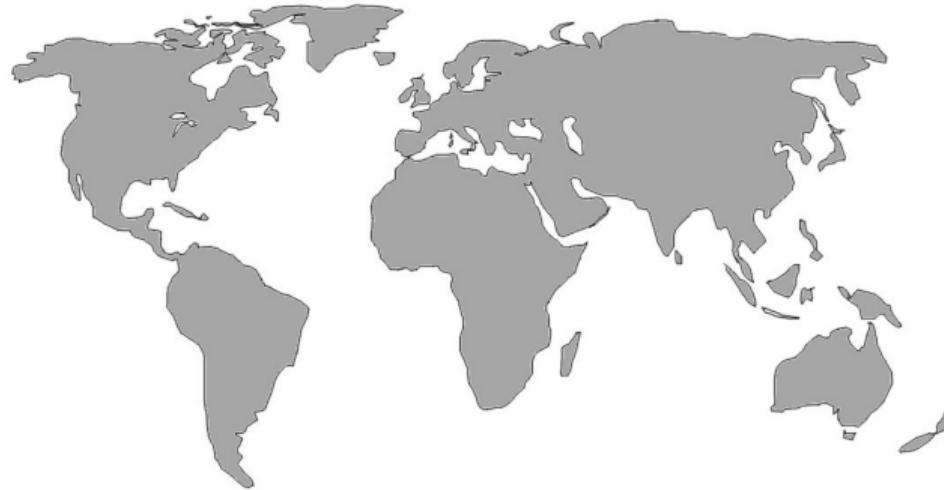


Scoring of shape and ESP similarity

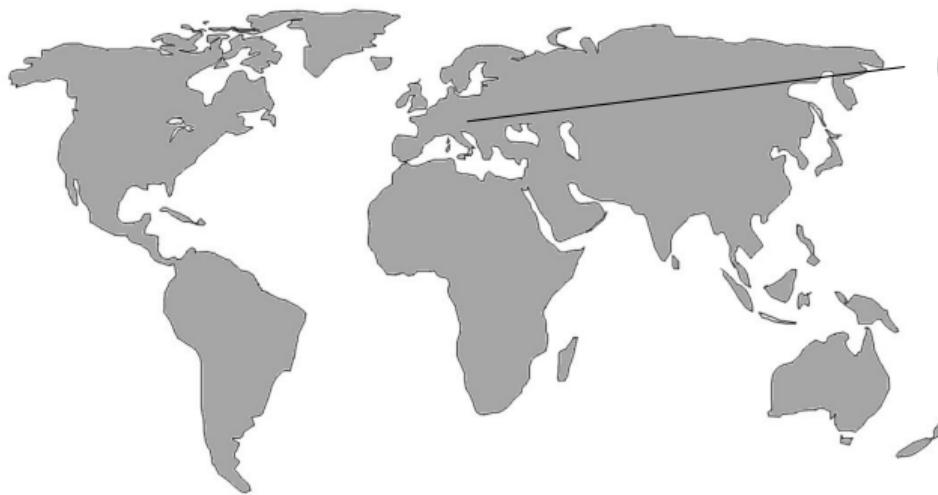
Dr. Esther Heid



About me

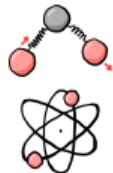


About me

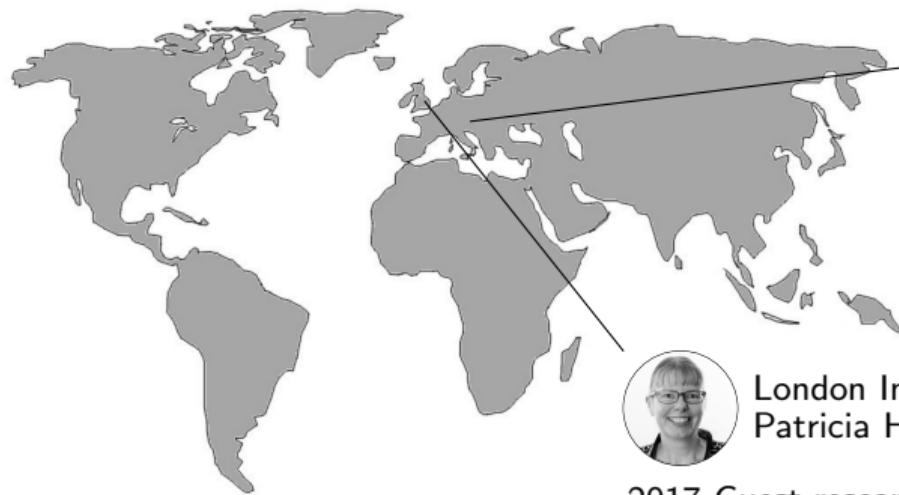


University of Vienna
Christian Schröder

2011 - 2014 BSc
2014 - 2016 MSc
2016 - 2019 PhD



About me



University of Vienna
Christian Schröder

2011 - 2014 BSc
2014 - 2016 MSc
2016 - 2019 PhD



London Imperial College
Patricia Hunt

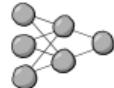
2017 Guest researcher



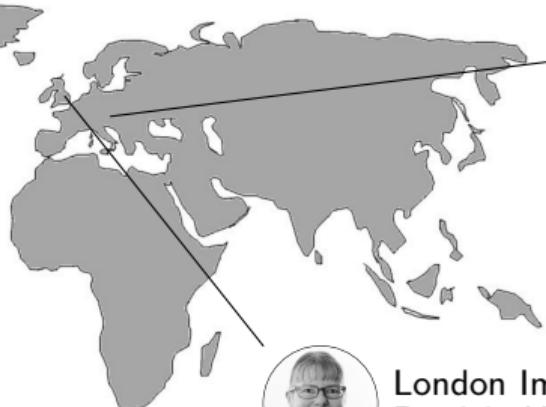
About me



2018 Guest researcher



University of Maryland
Alexander MacKerell



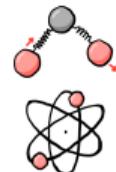
2017 Guest researcher



London Imperial College
Patricia Hunt



2011 - 2014 BSc
2014 - 2016 MSc
2016 - 2019 PhD



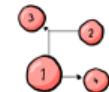
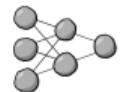
University of Vienna
Christian Schröder

About me



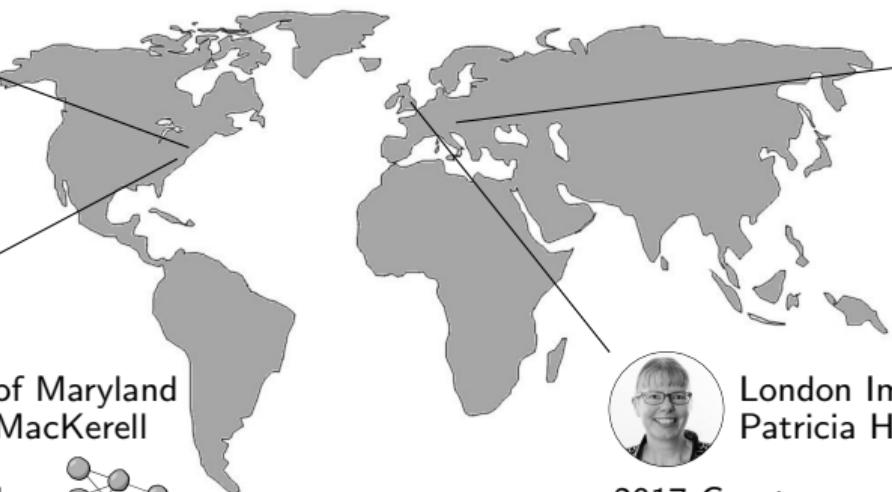
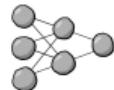
MIT
William Green

2020 - 2021 Postdoc



University of Maryland
Alexander MacKerell

2018 Guest researcher



University of Vienna
Christian Schröder

2011 - 2014 BSc

2014 - 2016 MSc

2016 - 2019 PhD



London Imperial College
Patricia Hunt

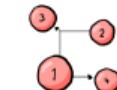
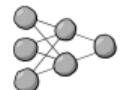
2017 Guest researcher



About me



MIT
William Green
2020 - 2021 Postdoc



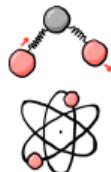
University of Maryland
Alexander MacKerell
2018 Guest researcher



TU Vienna
Georg Madsen
since 2022 Postdoc



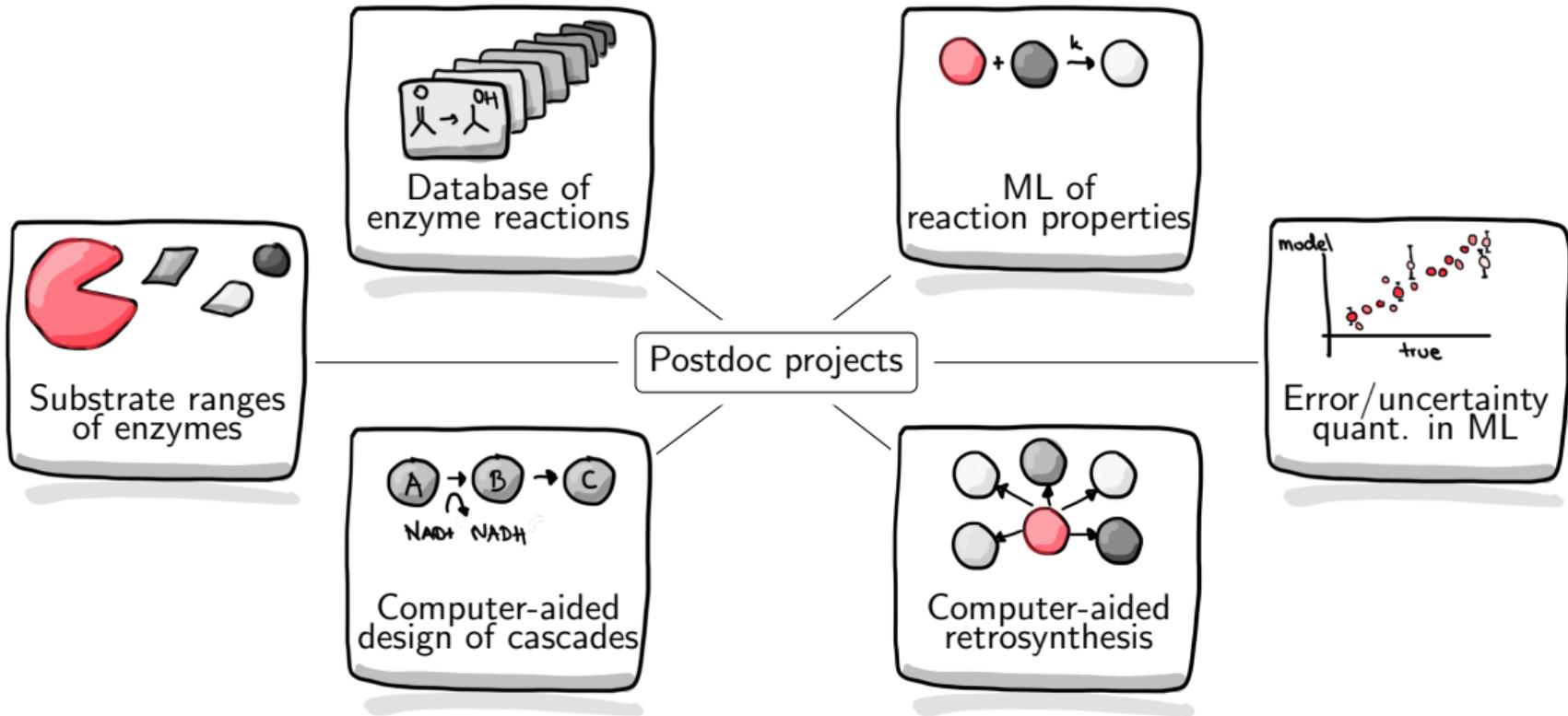
University of Vienna
Christian Schröder
2011 - 2014 BSc
2014 - 2016 MSc
2016 - 2019 PhD



London Imperial College
Patricia Hunt
2017 Guest researcher



Projects and interests



Projects and interests

Interests:

- Machine learning, deep learning
- Cheminformatics
- Reaction modeling
- Software development
- Computer simulation

Want to collaborate? Let me know:
esther.heid@tuwien.ac.at or eheid@mit.edu

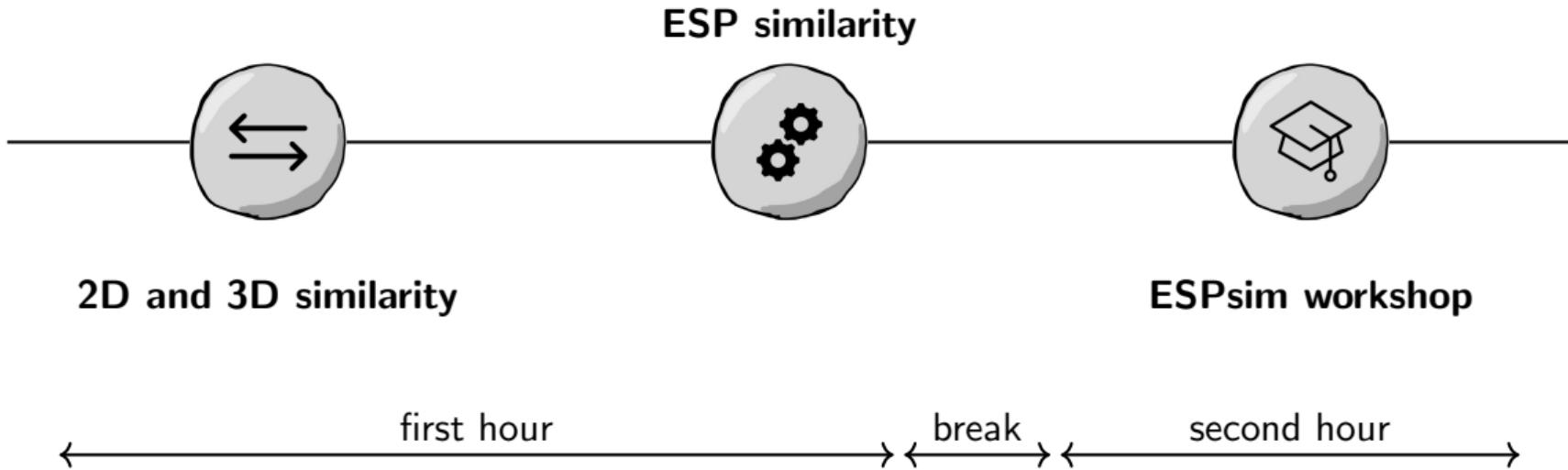


github.com/hesther



Google Scholar

Overview



Workshop materials

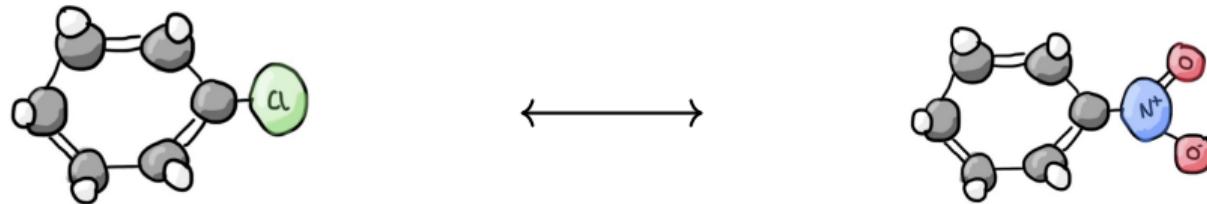
To follow the workshop (second hour), you can either use



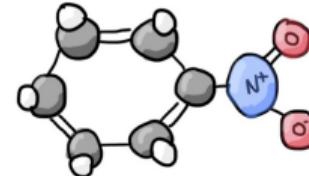
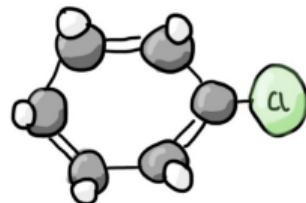
github.com/hester/espim/workshop

- Google Colab: No download or installation, but gmail account required
- Jupyter notebook: Installation required (pip or conda), follow instructions at github.com/hester/espim/workshop
- Simply listen in

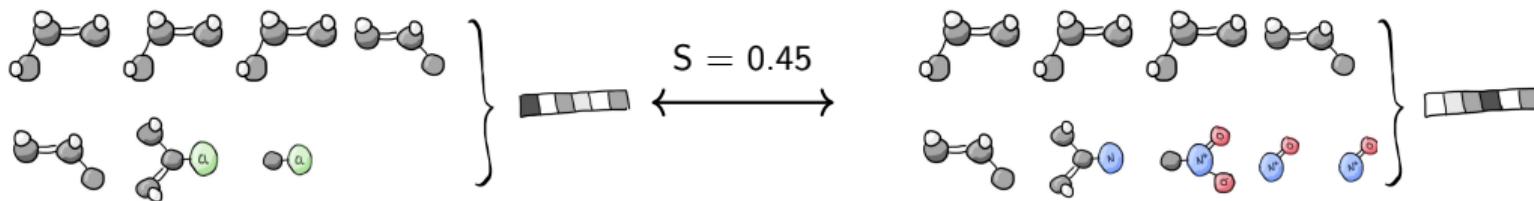
Similarity comparisons of molecules



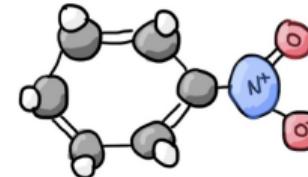
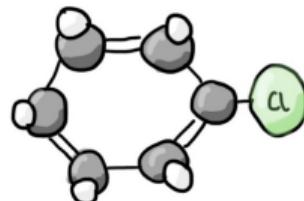
Similarity comparisons of molecules



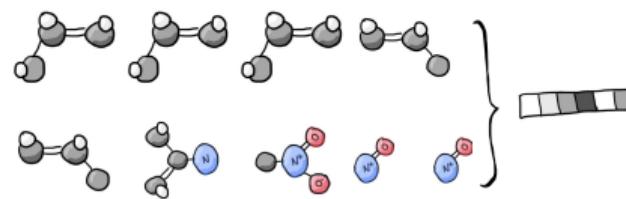
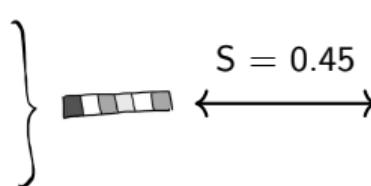
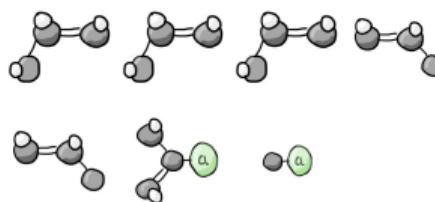
2D similarity:



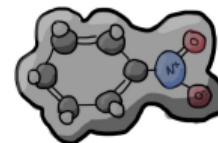
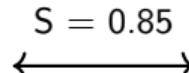
Similarity comparisons of molecules



2D similarity:

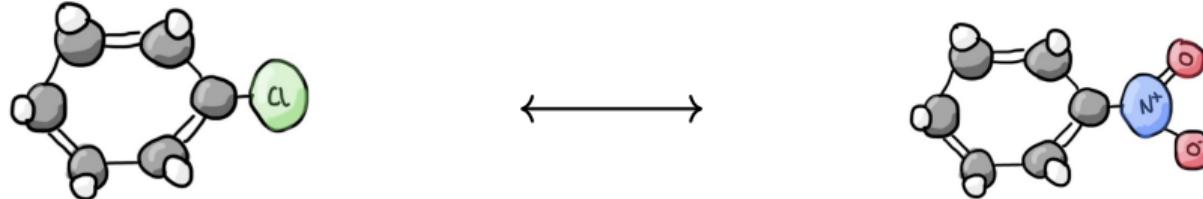


Shape similarity:

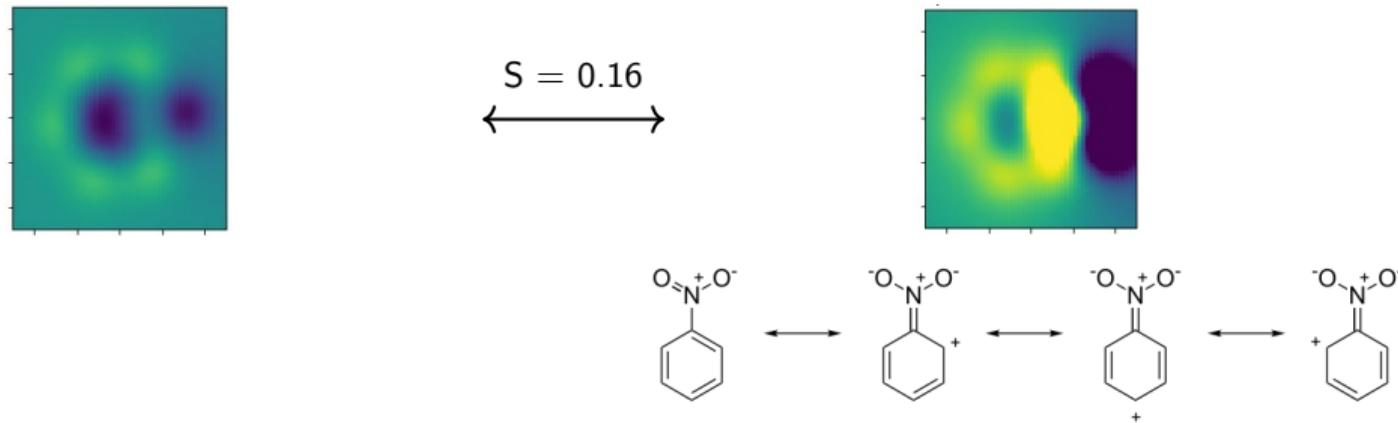


Similarity comparisons of molecules

Often neglected:

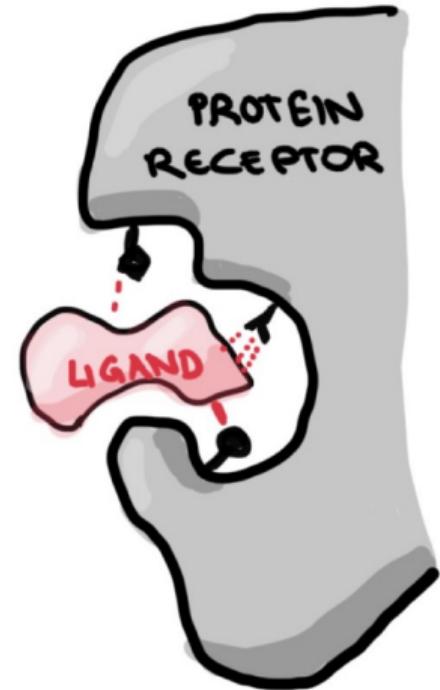
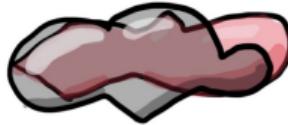


Electrostatic similarity:



Potential use cases of shape and ESP similarity

- Structure-activity relationships
- Protein-ligand scoring
 - compare known ligand to database of potential ligands
 - propose new ligands based on generative models
 - score enzymatic affinity towards new substrates
- Superimposing structures (by finding maximum similarity)

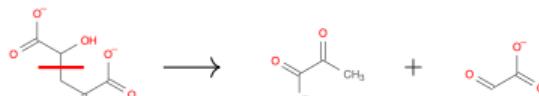


Substrate ranges of enzymes

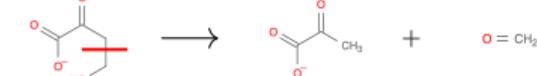
Example: (4S)-4-hydroxy-2-oxoglutarate lyase (EC 4.1.3.42)

Input:

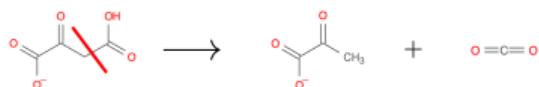
Reaction 1:



Reaction 2:



Reaction 3:

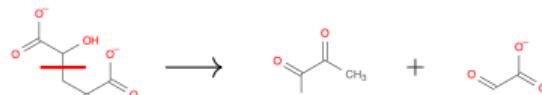


Substrate ranges of enzymes

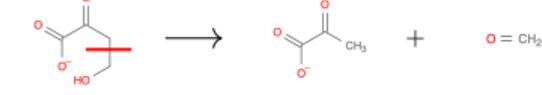
Example: (4S)-4-hydroxy-2-oxoglutarate lyase (EC 4.1.3.42)

Input:

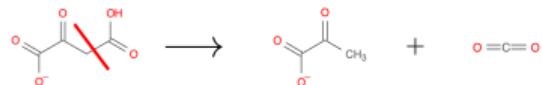
Reaction 1:



Reaction 2:



Reaction 3:



Standard templates:

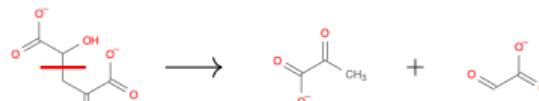


Substrate ranges of enzymes

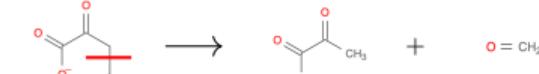
Example: (4S)-4-hydroxy-2-oxoglutarate lyase (EC 4.1.3.42)

Input:

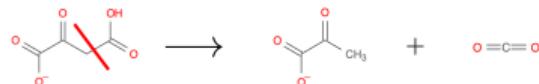
Reaction 1:



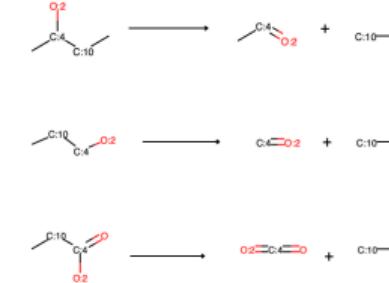
Reaction 2:



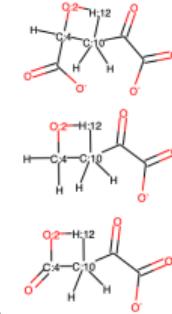
Reaction 3:



Standard templates:

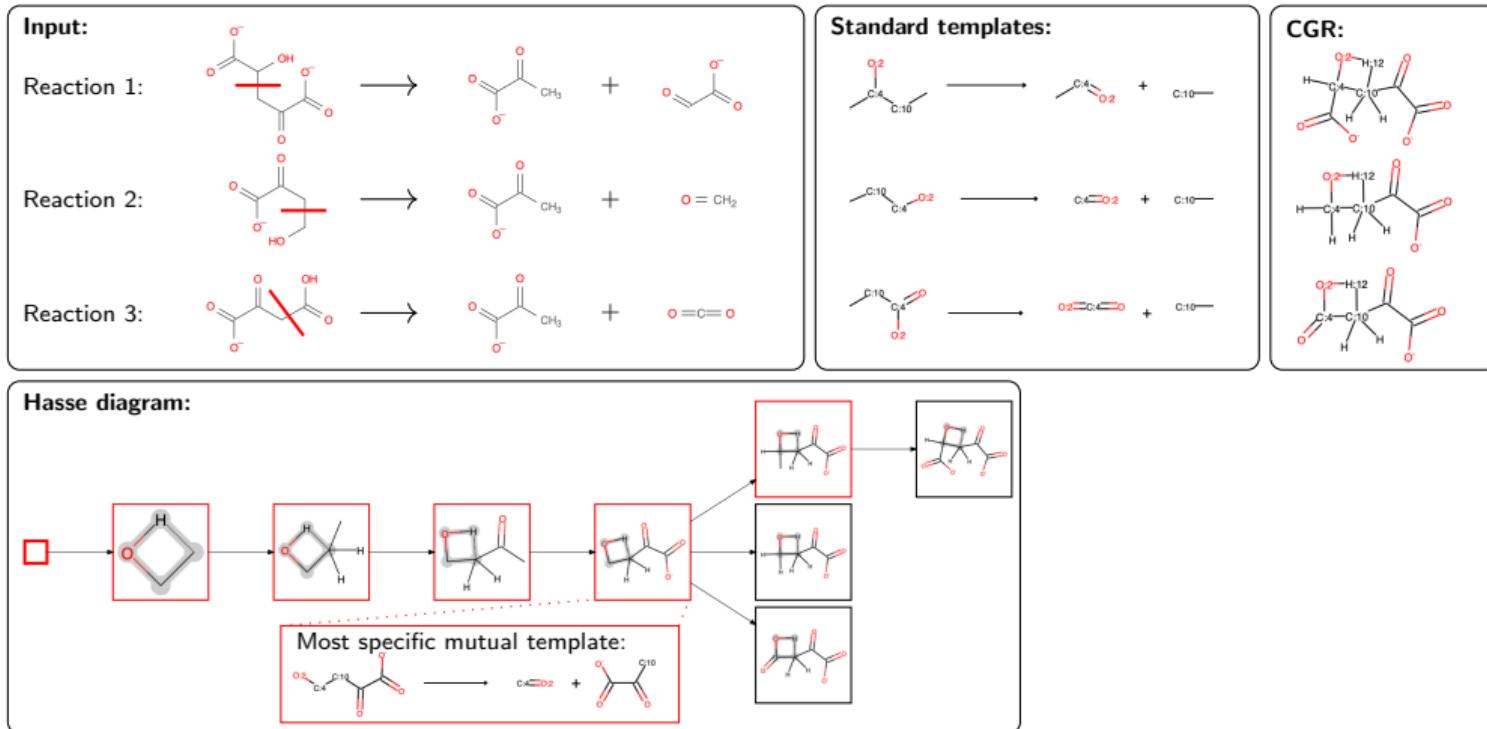


CGR:



Substrate ranges of enzymes

Example: (4S)-4-hydroxy-2-oxoglutarate lyase (EC 4.1.3.42)



Substrate ranges of enzymes

Example: (4S)-4-hydroxy-2-oxoglutarate lyase (EC 4.1.3.42)

Input:

Reaction 1:

Reaction 2:

Reaction 3:

Standard templates:

CGR:

Hasse diagram:

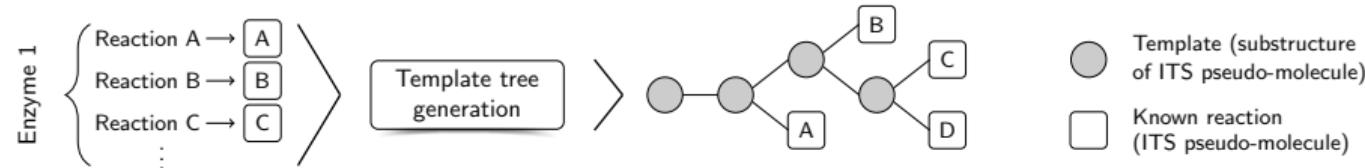
Fit into pocket

Riedel et al. PLoS One (2011) 6, e26021

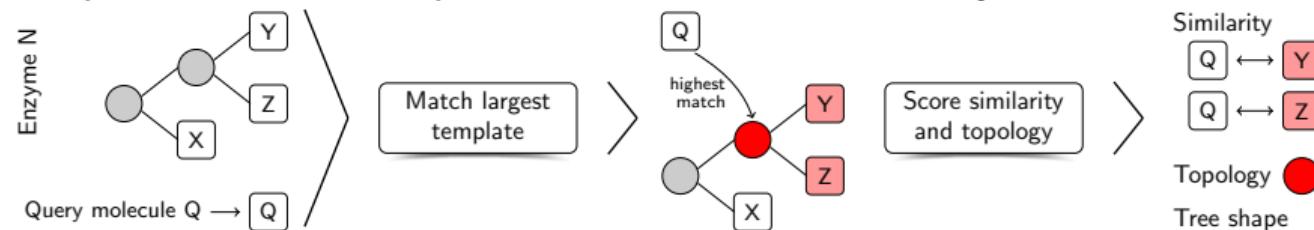
Substrate ranges of enzymes

EHreact: Extended Hasse diagrams for the extraction and scoring of reaction templates

- **Better reaction templates:** Extract reaction templates based on common substructures in the imaginary transition state (ITS), arrange in tree (Hasse diagram)



- **Better prediction of activity:** Based on the template tree shape (enzyme promiscuity), the position of the template in the tree, and similarity measures



Substrate ranges of enzymes

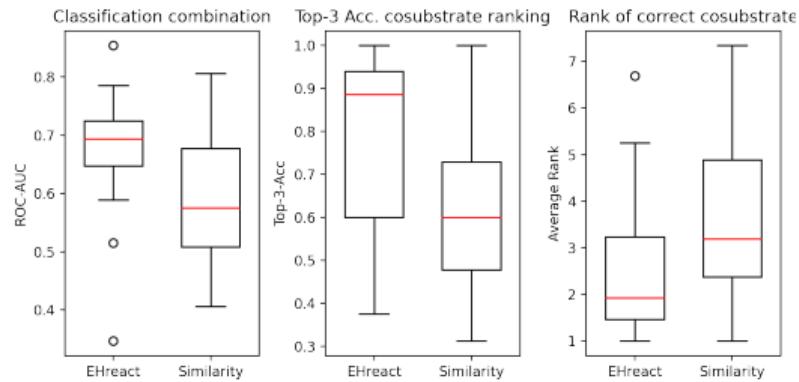
Scoring via Hasse diagram compared to scoring via molecular similarity:

Accuracy of classification (active/inactive substrates via leave-one-out):

	Sim	Hasse
Nitrilases	0.87	0.88
Aminedehydrogenases	0.89	0.90
Alcoholdehydrogenases	0.75	0.75
Carboxyl-methyltransferases	0.25	0.69
Transaminases	0.53	0.79
Tryptophansynthases	0.39	0.62
Amidinotransferases	0.19	0.76
Dehalogenases	0.53	0.75

TODO: Add 3D similarity

Cosubstrate proposal (for 18 transaminase reactions):



E. Heid et al. *J. Chem. Inf. Model* 2021, 61, 4949.
Code: <https://github.com/hesther/ehreact>

Available options for shape and electrostatic similarity

OpenEye:

- ROCS (shape similarity)
- EON (electrostatic similarity) via Poisson-Boltzmann solver or Coulombic model*

Surflex:

- eSim (grid-based 3D molecular similarity via electrostatic field, molecular surface-shape, directional hydrogen-bonding)

AMD tools:

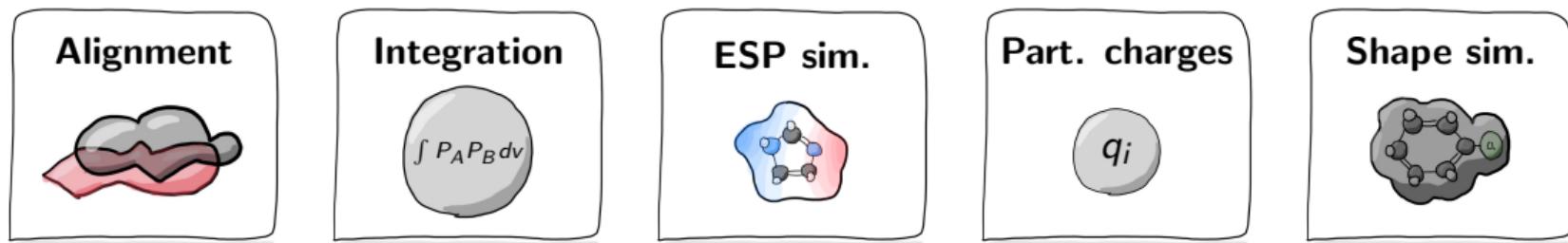
- mRAISE (descriptor-based bitmap search engine based on shape-scores)

No open-source options which are freely available independent of the use case

* which is often equally good to judge the similarity between molecules, see
Jennings et al, *J. Chem. Inf. Model* 2007, 47, 1829

ESPsim

Easy-to-use python package built on RDKit, fully open-source, freely available



Constrained and unconstrained alignment via RDKit

Analytic integration via Gaussians + numeric MC integration

Carbo or Tanimoto similarity via Coulomb

Bolcato, Heid, Bostrom, *J. Chem. Inf. Model.* 2022, 62, 1388

github.com/hesther/espSim

Comparing electrostatic potentials



Electron cloud: Could depict electronegativities, inductive and mesomeric effects

But: Electron clouds take time to compute, instead use ESPs

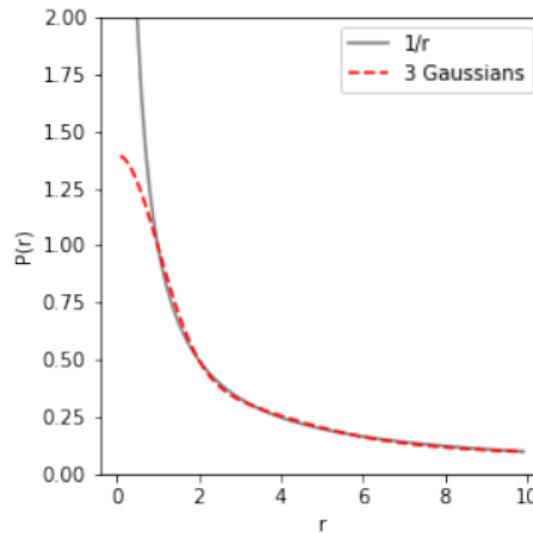
Compare electrostatic potentials via Coulomb potential of trial charge at r : $P(r) = \sum_i \frac{q_i}{r - R_i}$

Carbo et al, *Int. J. Quantum Chem.* 1980, 17, 1185
Good et al, *J. Chem. Inf. Comput. Sci.* 1992, 32, 188

Comparing electrostatic potentials

But: $\lim_{r \rightarrow 0} \frac{1}{r} = \infty$

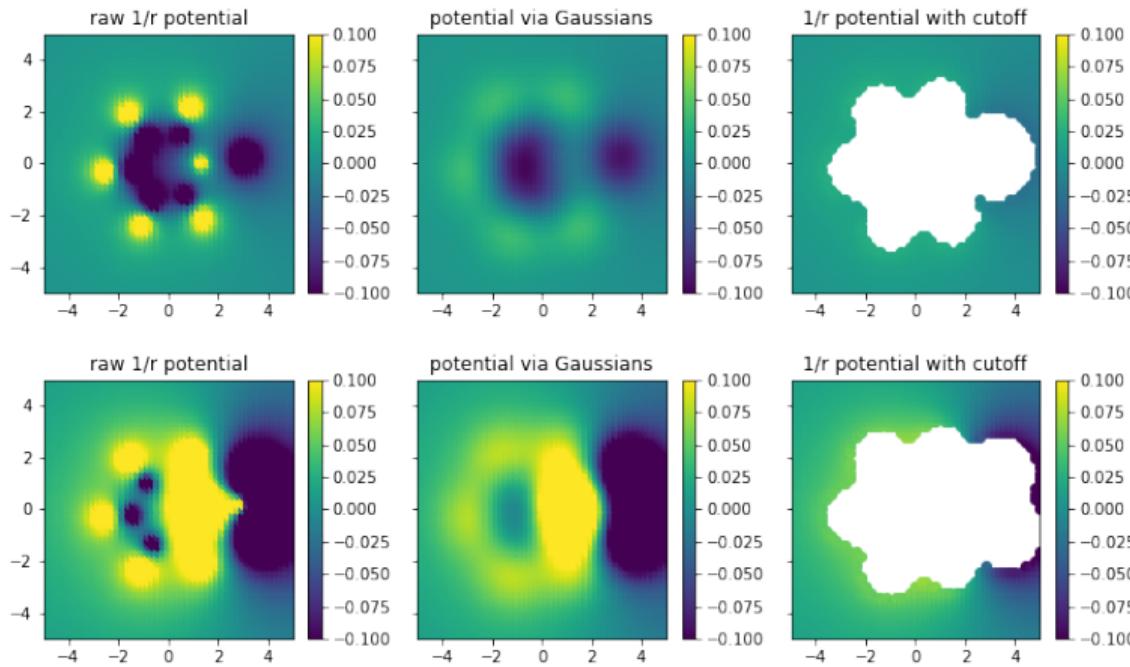
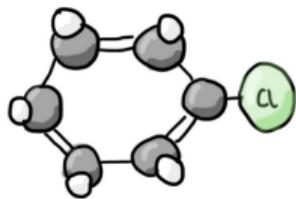
- comparison is only valid outside of the atomic nuclei (usually vdW radii), evaluate on a grid
- fit $\frac{1}{r}$ with Gaussians



$$P(r) = \sum_i \frac{q_i}{r - R_i} \simeq \sum_i q_i (0.3e^{-0.05(r-R_i)^2} + 1.0e^{-0.5(r-R_i)^2} + 0.1e^{-0.003(r-R_i)^2})$$

Good et al, *J. Chem. Inf. Comput. Sci.* **1992**, 32, 188

Comparing electrostatic potentials

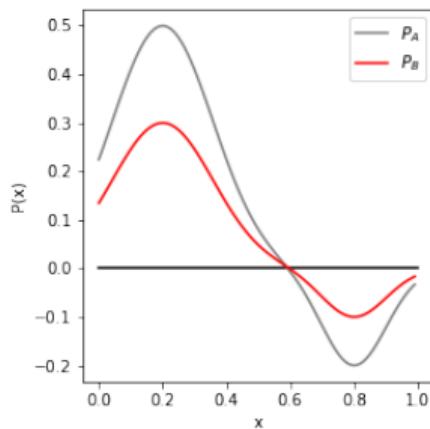


Comparing electrostatic potentials

Quantify similarity of ESPs via overlap of functions $\int P_A P_B dv$:

- **Carbo similarity** $S = \frac{\int P_A P_B dv}{\sqrt{\int P_A^2 dv} \sqrt{\int P_B^2 dv}}$ range -1 to 1
- **Tanimoto similarity** $S = \frac{\int P_A P_B dv}{\int P_A^2 dv + \int P_B^2 dv - \int P_A P_B dv}$ range -1/3 to 1

Difference:

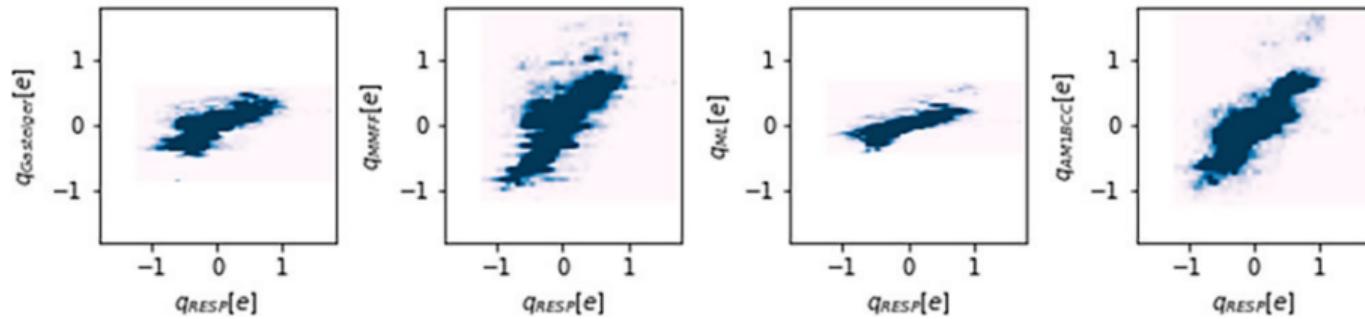


- $S_{Carbo} = 0.999$ (insensitive to magnitude)
- $S_{Tanimoto} = 0.778$ (sensitive to magnitude)

The influence of partial charges

Multitude of partial charge distributions available!

Comparison of ESPs from QM RESP charges to non-QM partial charge distributions:

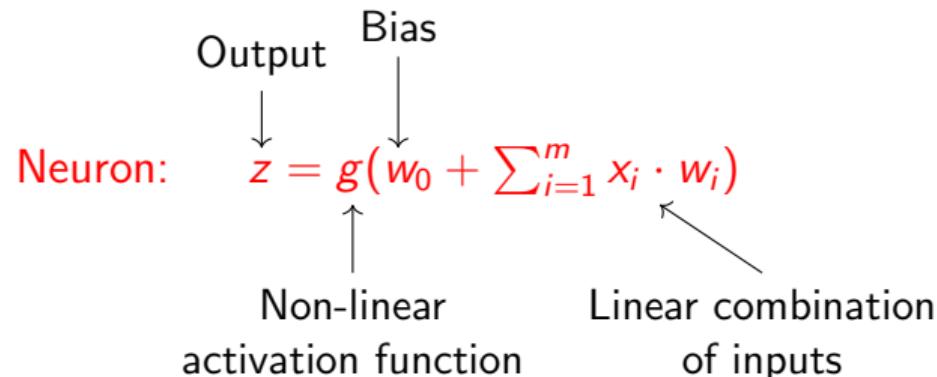
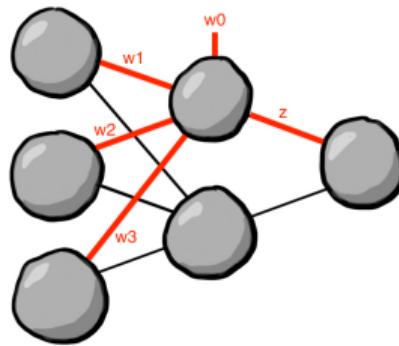


	MAE [e]	S_{Carbo}	$S_{Tanimoto}$
Gasteiger (RDKit)	0.16	0.78	0.60
MMFF94	0.17	0.80	0.64
ML model	0.17	0.85	0.61
AM1-BCC	0.12	0.88	0.78

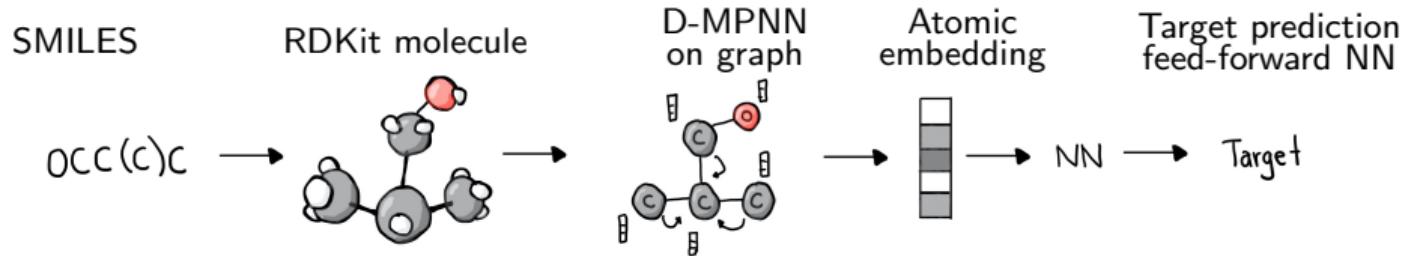
Bolcato, Heid, Bostrom, *J. Chem. Inf. Model.* 2022, 62, 1388
Guan, Coley, Wu, Ranasinghe, Heid, Struble, Pattanaik, Green, Jensen *Chem. Sci.* 2021, 12, 2198

Machine-learned partial charges

Neural networks:

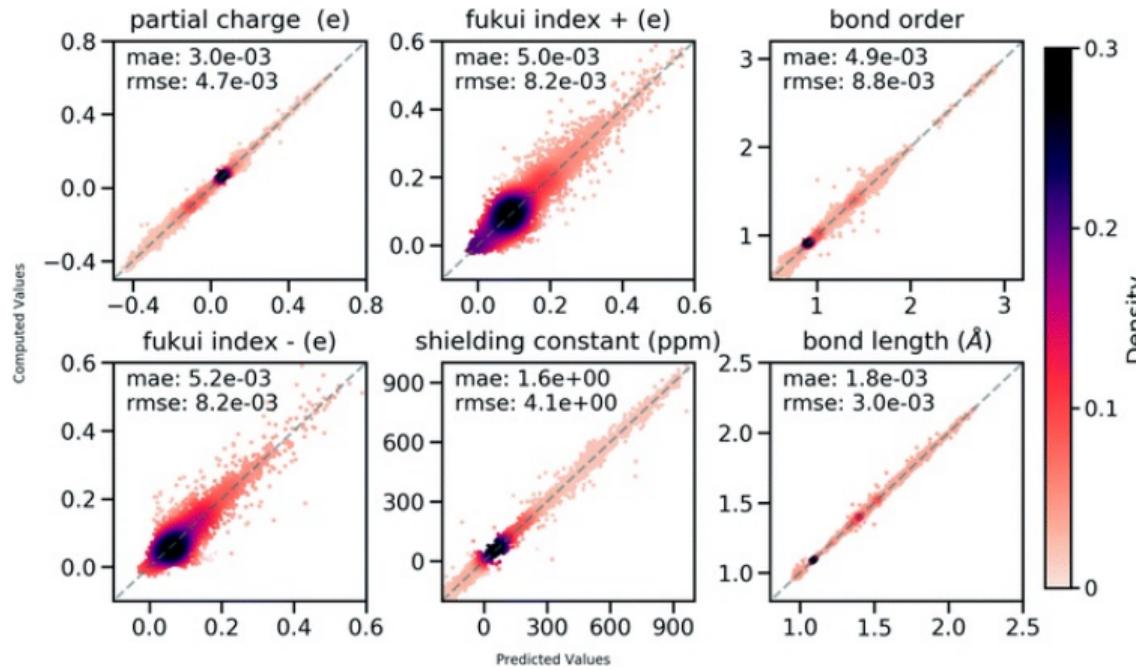


Directed message-passing neural networks for chemistry applications:



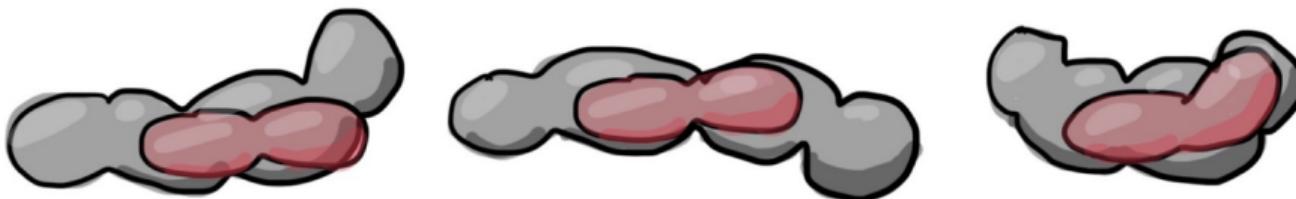
Machine-learned partial charges

D-MPNN neural network trained on QM Hirshfeld charges of 136k molecules (4.4M atoms)



Guan, Coley, Wu, Ranasinghe, Heid, Struble, Pattanaik, Green, Jensen *Chem. Sci.* **2021**, 12, 2198

Aligning molecules: Unconstrained



Alignment of any pair of molecules:

- Random embedding of N conformations
- Alignment via Open3DAlign with mapping based on Crippen logP atom contributions (as implemented in RDKit)
- Choose best conformers based on shape or ESP similarity

Requirements: None

Tosco et al *J. Comput-Aid. Mol. Design* 2011, 25, 777

Aligning molecules: Constrained



Alignment with constrained core, useful for e.g. enzymatic reactions

- Constrained embedding of N conformations with pre-defined coordinates of core
- Alignment via Open3DAlign with mapping based on common core (as implemented in RDKit)
- Choose best conformers based on shape or ESP similarity

Requirements: Large common core of molecules, e.g. via EHreact software

Aligning molecules: Custom



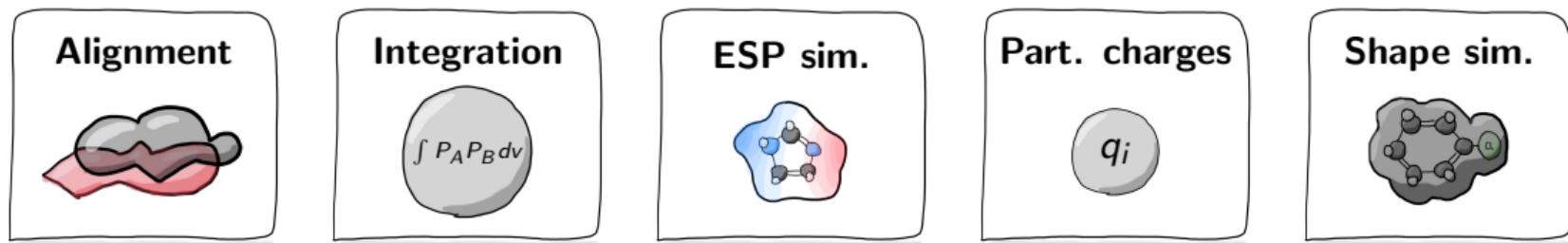
Use pre-made alignments, for example from docking:

- Load aligned molecules
- Compute single value of shape or ESP similarity

Requirements: Known aligned conformations

ESPsim

Easy-to-use python package built on RDKit, fully open-source, freely available



Constrained and unconstrained alignment via RDKit
Analytic integration via Gaussians + numeric MC integration
Carbo or Tanimoto similarity via Coulomb

Bolcato, Heid, Bostrom, *J. Chem. Inf. Model.* 2022, 62, 1388

github.com/hesther/espssim

Benchmark: Binding to D4 dopamine receptor

Experimental study of 549 compounds for interactions with D4

Lyu et al, *Nature* 2019, 566, 224

Re-scoring via different algorithms:

		AUC	BEDROC
ESPsim	Tanimoto (this work)	0.63	0.43
	Smina Score	0.31	0.31
	RFScore	0.42	0.19
	PLECSore	0.58	0.28
	NNScore	0.50	0.22
	Feature Map Vector Score	0.65	0.44
	RF Score VS	0.51	0.18

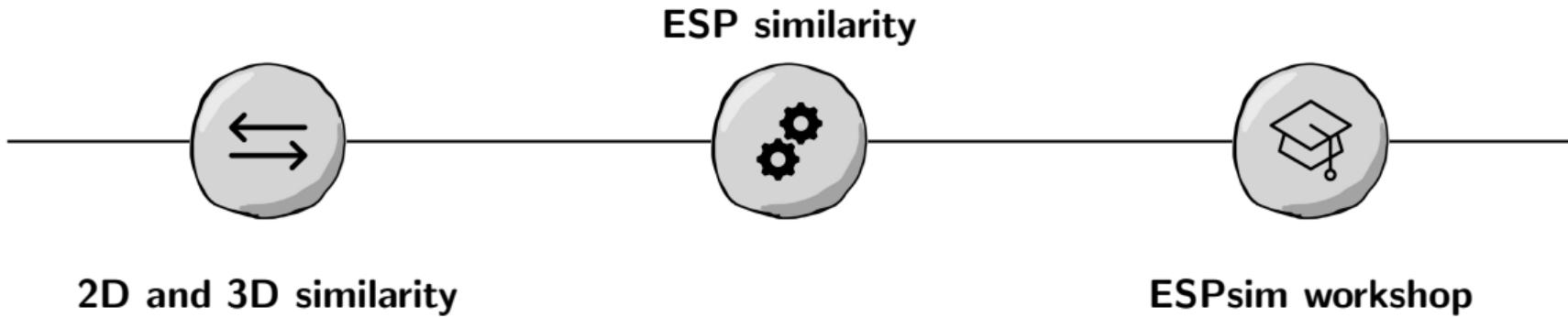
github.com/ljmartin/d4-rescore

Benchmark: DUD-E targets

Re-scoring of 102 DUD-E targets:

	AUC
ESPsim Tanimoto (this work)	0.70
mRAISE	0.74
eSim full	0.76
eSim only Shape + Coulomb	0.67
ROCS Color	0.66
ROCS Shape	0.60
WEGA	0.56
OptiPharm	0.56
VAMS	0.56
USR	0.55

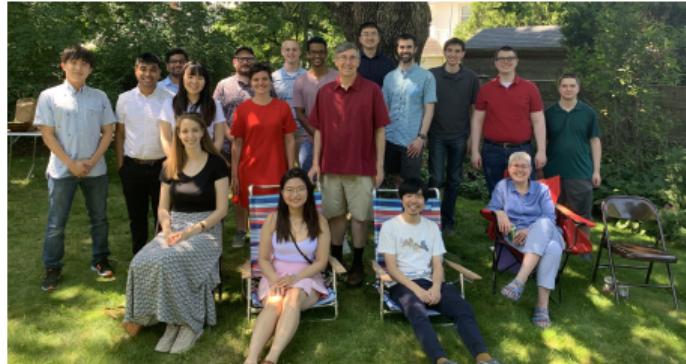
Summary



Acknowledgements



Prof. Green & the Green group



universität
wien Prof. Christoph Flamm



Prof. Jensen & the Jensen group



Prof. Madsen & the Madsen group

Thank you for your attention!
Please contribute on Github!



github.com/hester



Google Scholar

Funding:



Workshop materials

Part 2: Please follow the QR-code/link and use



github.com/hester/espim/workshop

- Google Colab: No download or installation, but gmail account required
- Jupyter notebook: Installation required (pip or conda), follow instructions at github.com/hester/espim/workshop
- Simply listen in