HemeSchNet

GöAID Seminar series

J. Jones, 02.06.2026

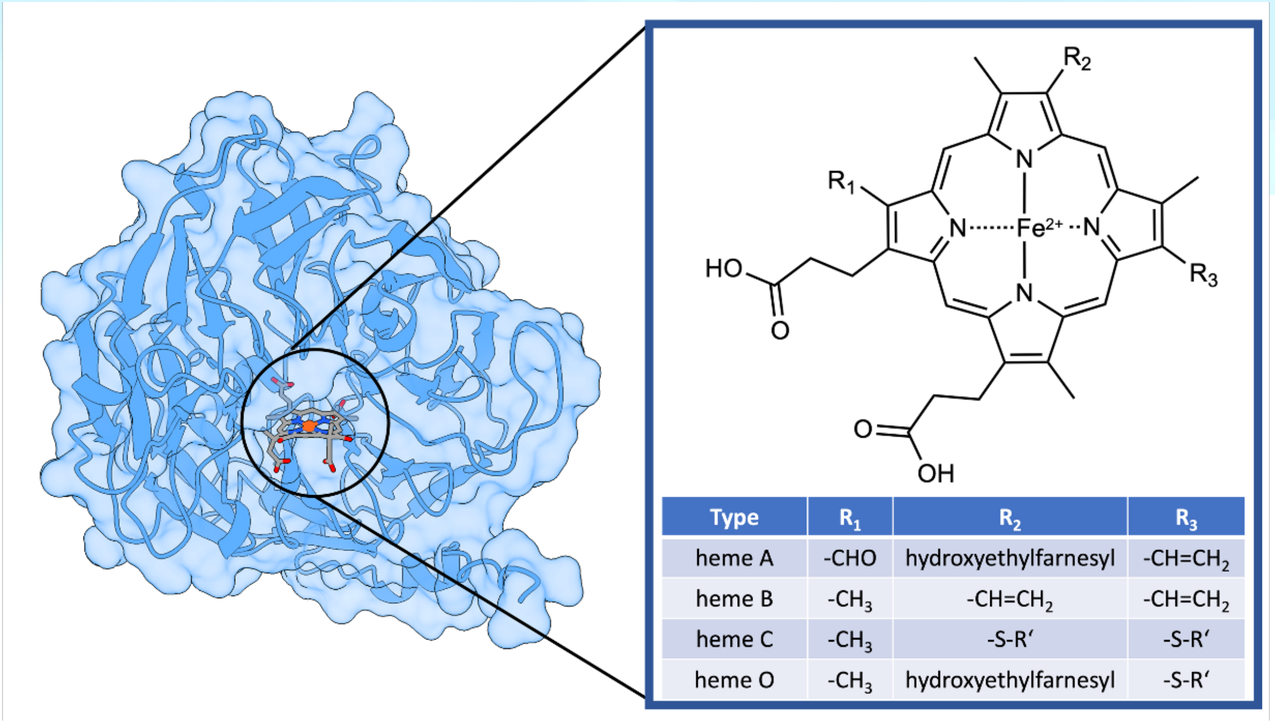
Talk Outline

- Problem Setting:
  - Heme Basics
  - Heme Distortions
- The Heme Electronic Structure Dataset HESD
- SchNet
- HemeSchNet

Heme Basics

Problem Setting

- Heme is a cofactor in enzymes, often the active centre
- >13.000 resolved structures
- Involved in oxygen transport and reduction, electron and proton transport, transcriptional regulation, etc.
- Iron-ion coordinated by 4 porphyrin ring nitrogens and ≤ 2 axial ligands



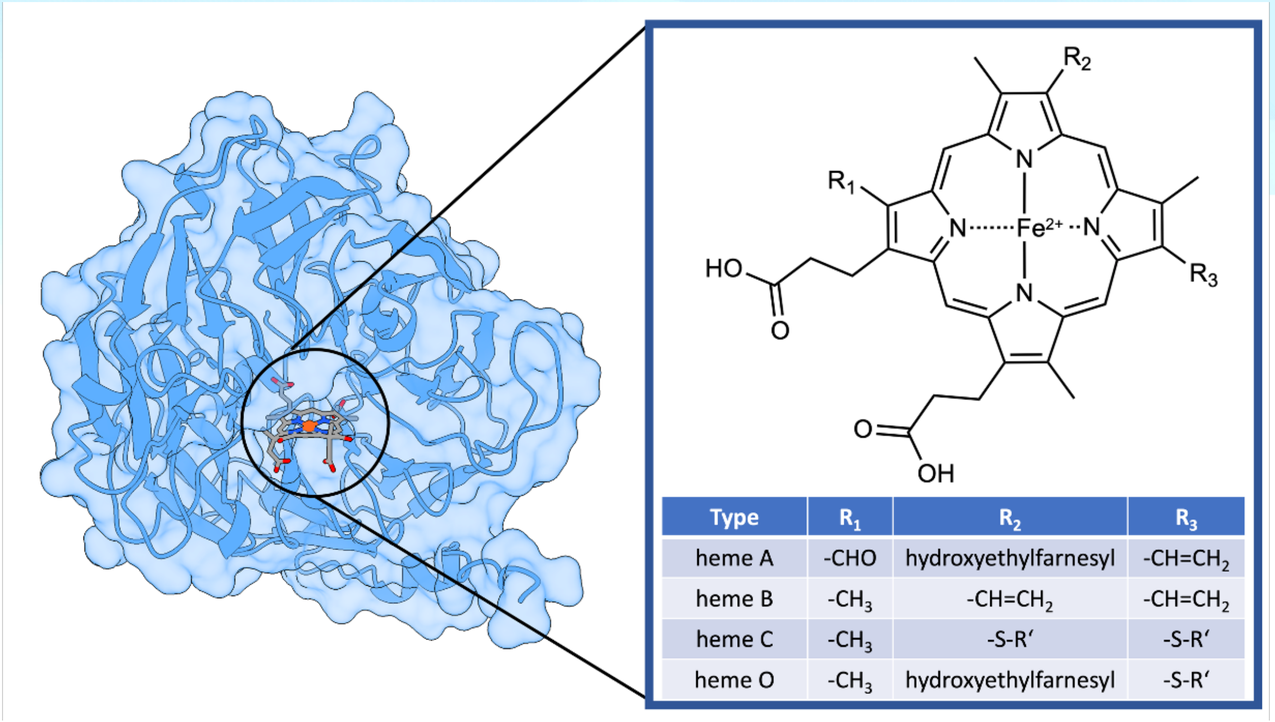
Type	R_1	R_2	R_3
heme A	-CHO	hydroxyethylfarnesyl	$-CH=CH_2$
heme B	$-CH_3$	$-CH=CH_2$	$-CH=CH_2$
heme C	$-CH_3$	-S-R‘	-S-R‘
heme O	$-CH_3$	hydroxyethylfarnesyl	-S-R‘

Heme Basics

Problem Setting

Chemical diversity is small, but functional diversity is huge - heme structure-function relationship is an open question

Assumption: Heme Function is somehow (electrostatics, QM) encoded in atomic positions

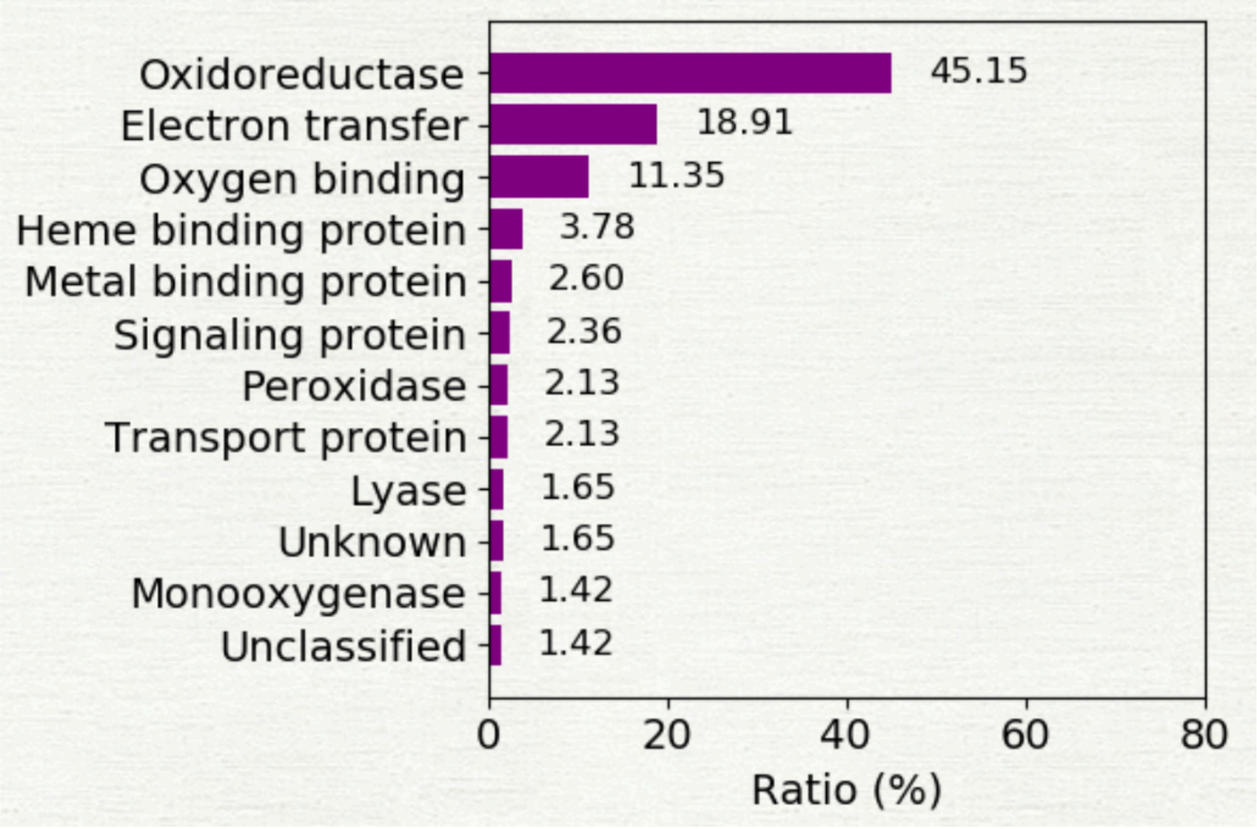


Type	R_1	R_2	R_3
heme A	-CHO	hydroxyethylfarnesyl	$-CH=CH_2$
heme B	$-CH_3$	$-CH=CH_2$	$-CH=CH_2$
heme C	$-CH_3$	-S-R‘	-S-R‘
heme O	$-CH_3$	hydroxyethylfarnesyl	-S-R‘

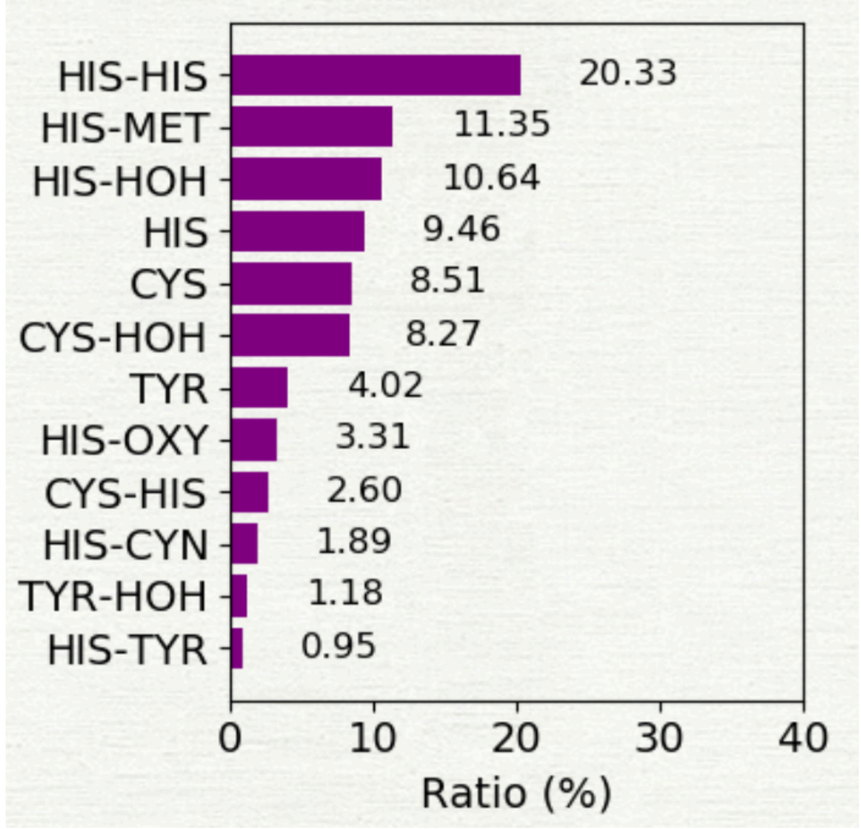
Heme(s) in Proteins

16.830 structures of hemes in proteins

Python-based database & analyzer for DIStortion of Heme porphyrin



Protein Function in pyDISH

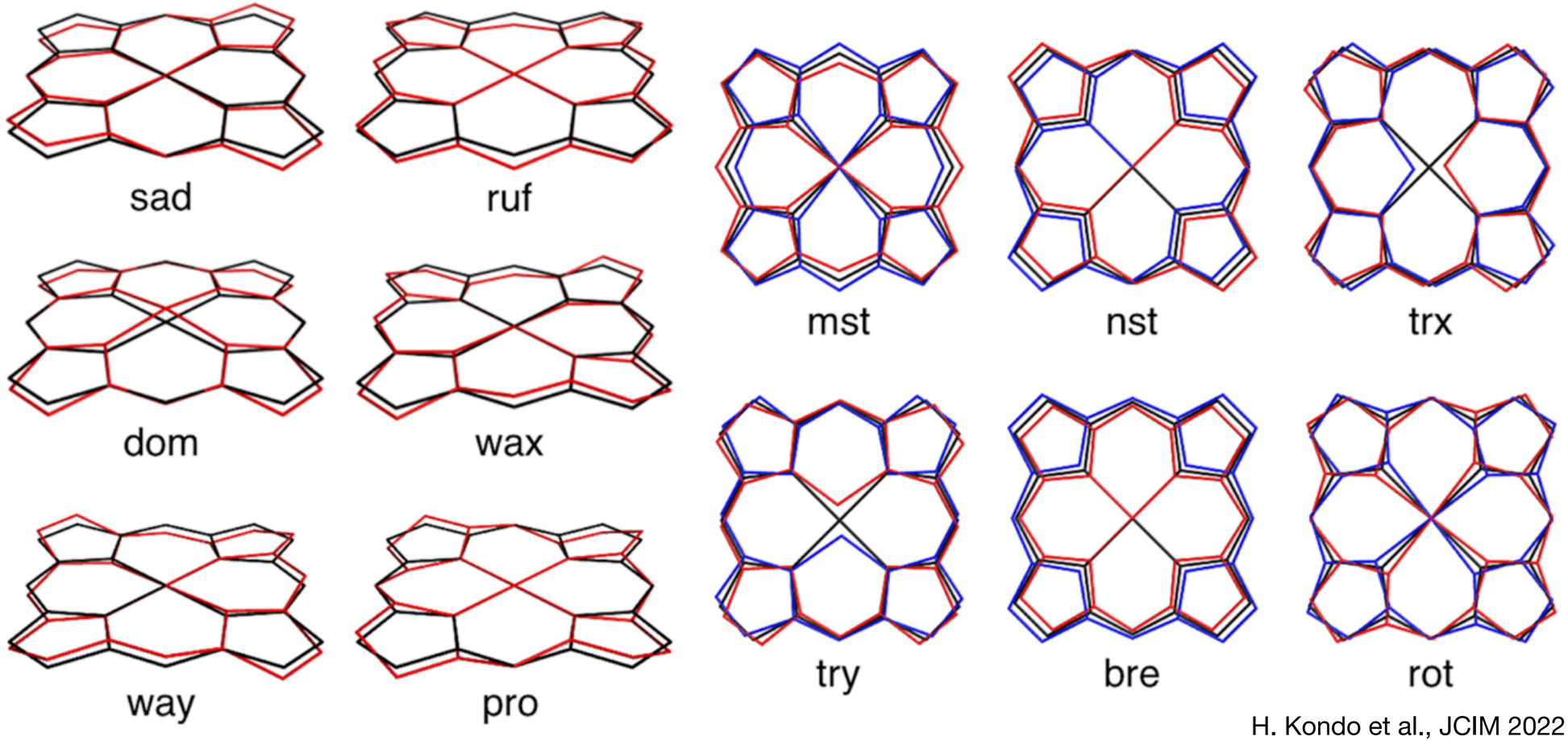


Axial Ligand Occurrence in pyDISH

H. Kondo et al., Database 2020

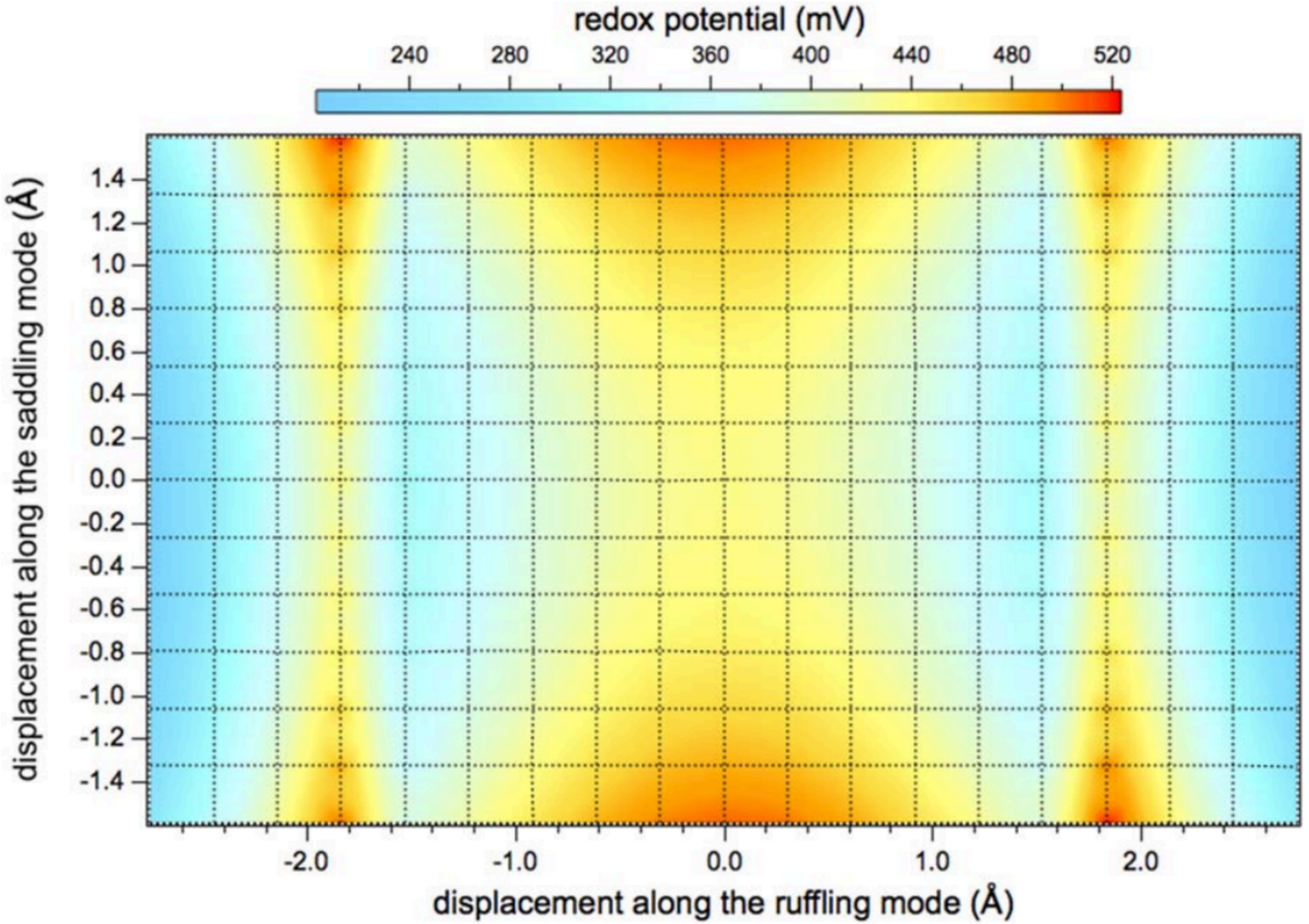
Heme Distortions

Problem Setting - Distortions



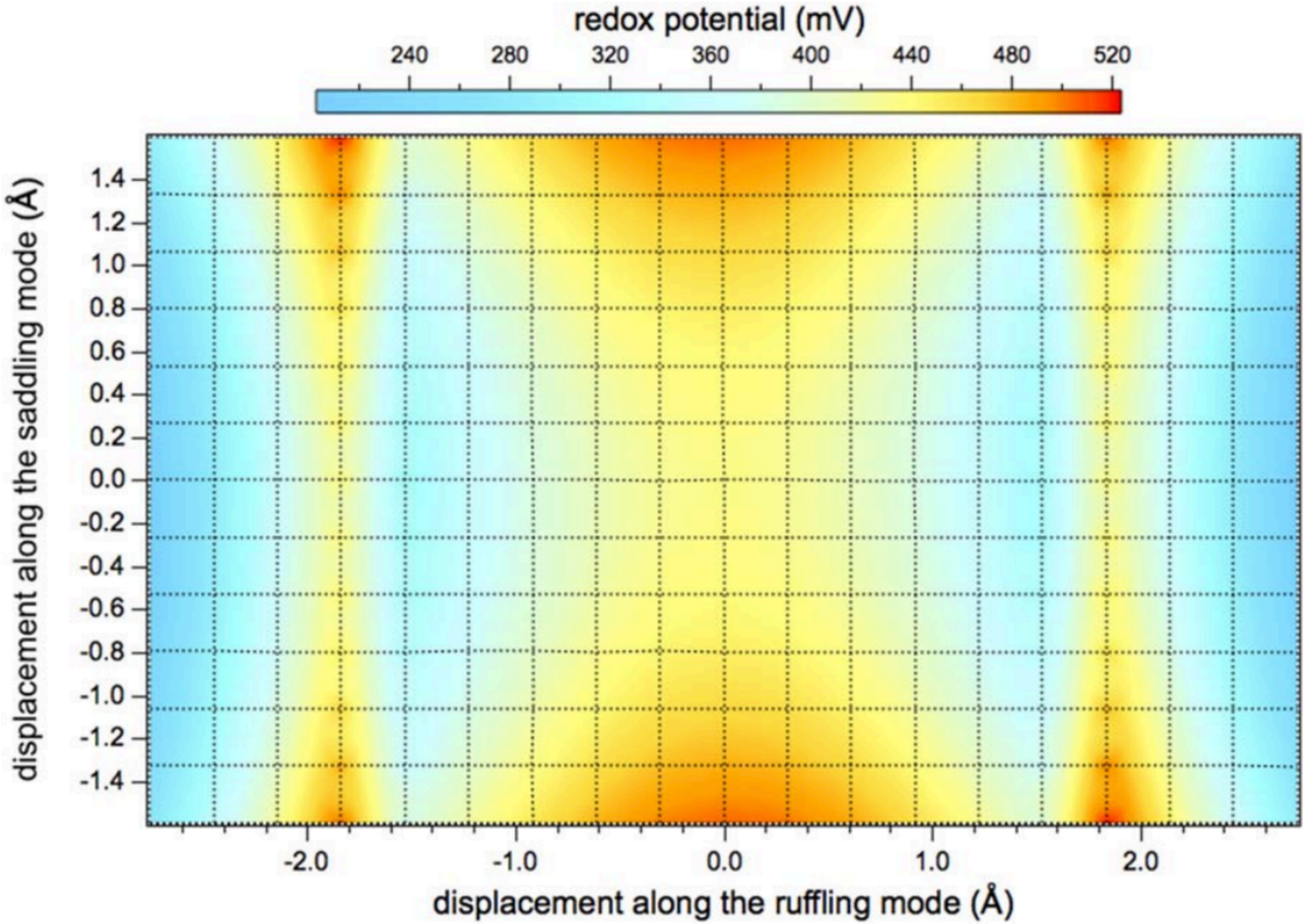
H. Kondo et al., JCIM 2022

Heme Distortions



Y. Takano et al., JCC 2018

Heme Distortions



There is a correlation between heme distortion and redox potential, but it is complex

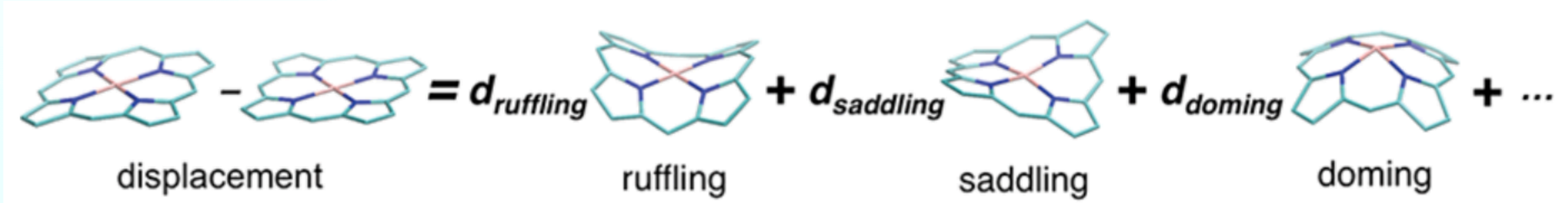
Y. Takano et al., JCC 2018

pyDISH

Python-based database & analyzer for DIStortion of Heme porphyrin

H. Kondo et al., Database 2020

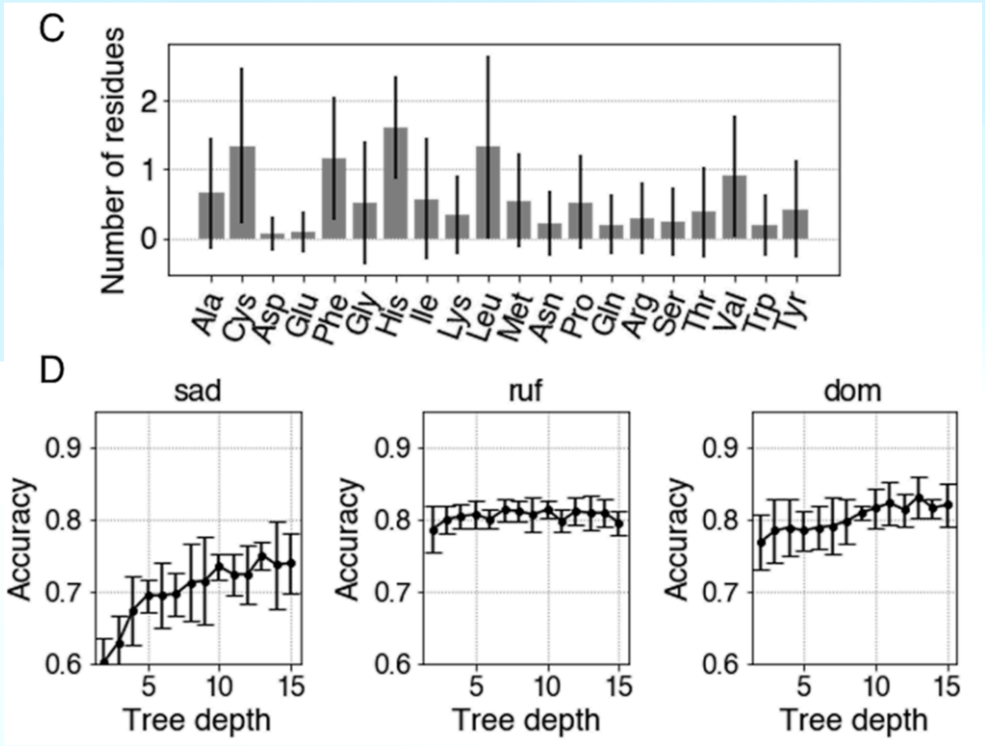
- pyDISH (https://pydish.bio.info.hiroshima-cu.ac.jp/) is a database where for all available heme crystal structures the Distortions are calculated and stored
  - 20.147 heme structures from 7441 PDB structures, regularly updated
- Distortions (displacement differences) are calculated based on normal-coordinate structure decomposition of the linear combination of different normal modes



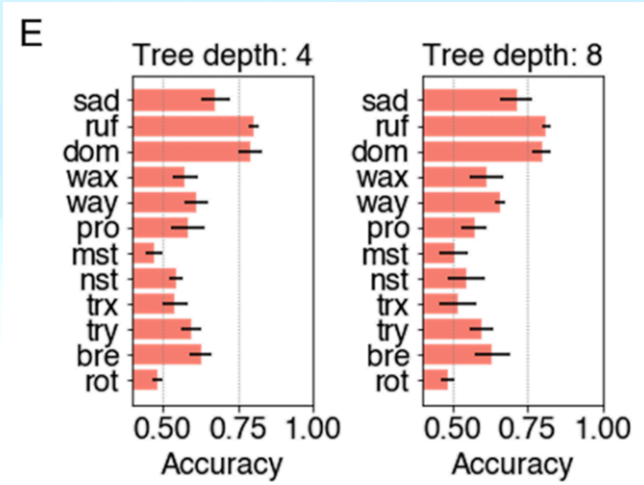
Y. Takano et al., JCC 2018

Binding Pocket Analysis

Occurrence of Residues in Heme Binding Pockets



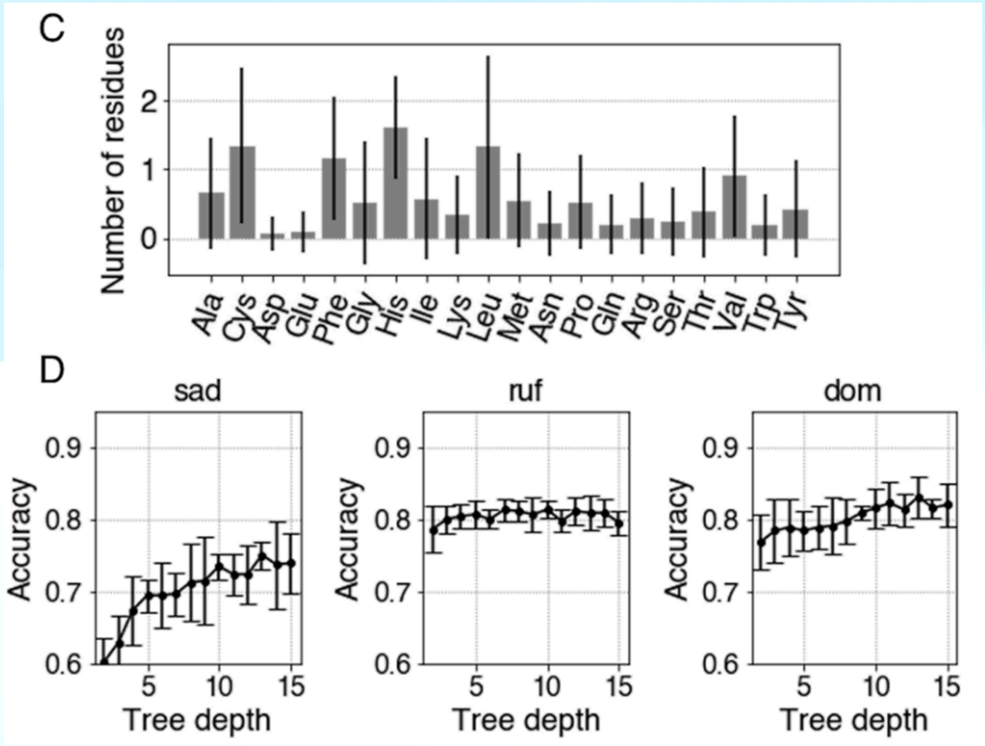
Accuracy of top three distortions based on tree depth:
Saddling (sad), Ruffling (ruf), Doming (dom)



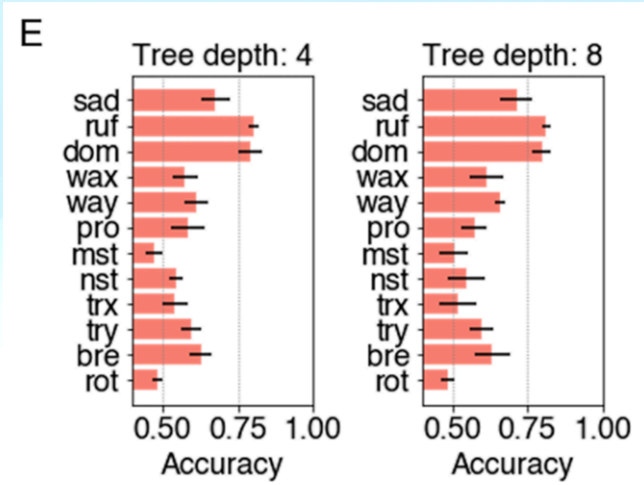
Mean Accuracy of Distortions obtained with
5-fold Cross-Validation

Binding Pocket Analysis

Occurrence of Residues in Heme Binding Pockets



Accuracy of top three distortions based on tree depth: Saddling (sad), Ruffling (ruf), Doming (dom)



Mean Accuracy of Distortions obtained with 5-fold Cross-Validation

There is a correlation between binding pocket and heme distortion, but it is complex

The Heme Electronic Structure Dataset HESD

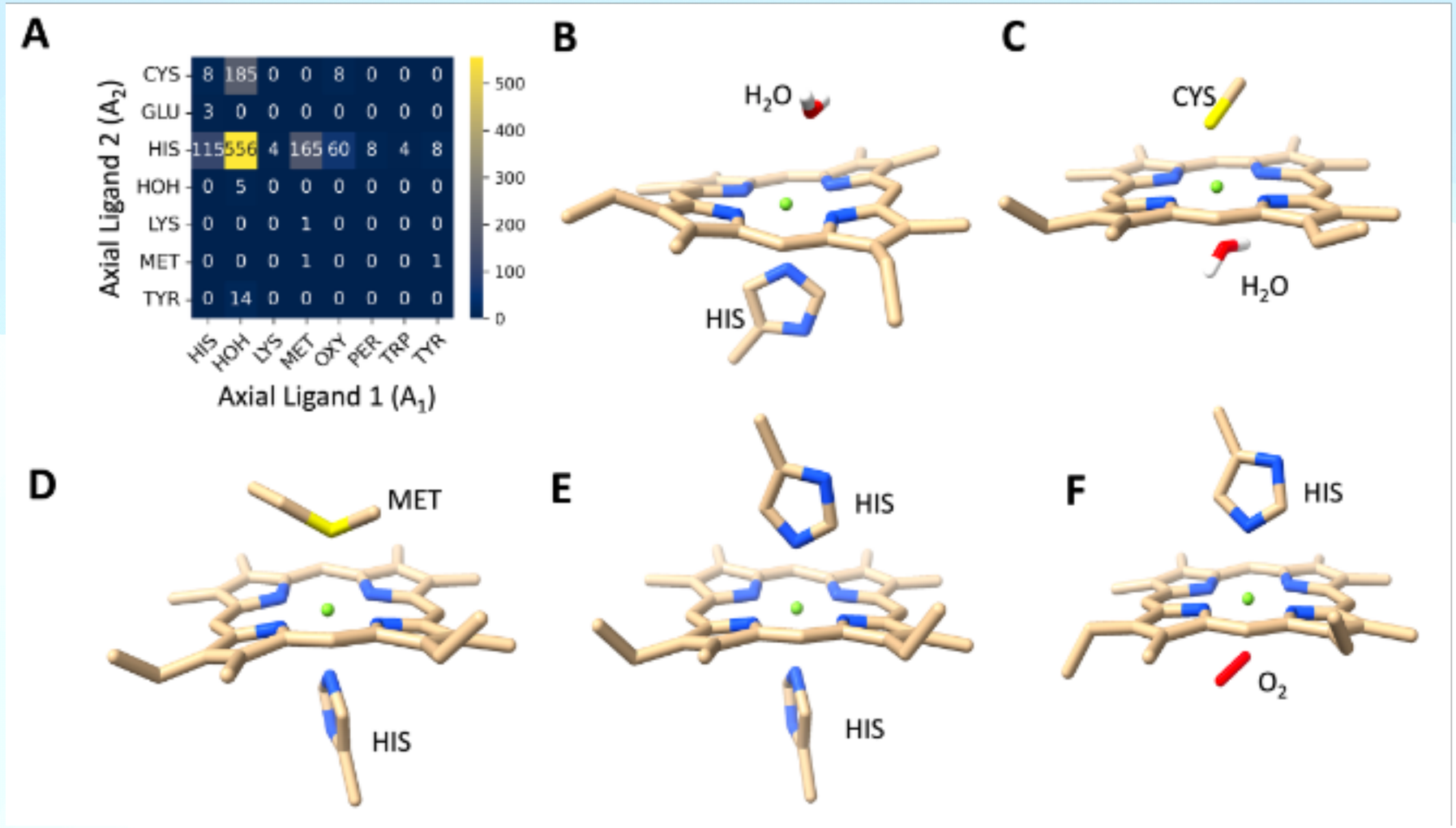
(Jones et al., 2026, In Review)

- We collected literature values on experimental redox potentials of hemes
- We could not reproduce any meaningful correlation between redox potential and distortions or redox potential and enzyme reaction family
- So we designed a computing scheme with which to capture the hemes electronic structure and calculated a database for it:
  - Models include Heme Porphyrin and axial ligands' side chains
  - Calculate Fe^{2+} and Fe^{3+} low/high spin models

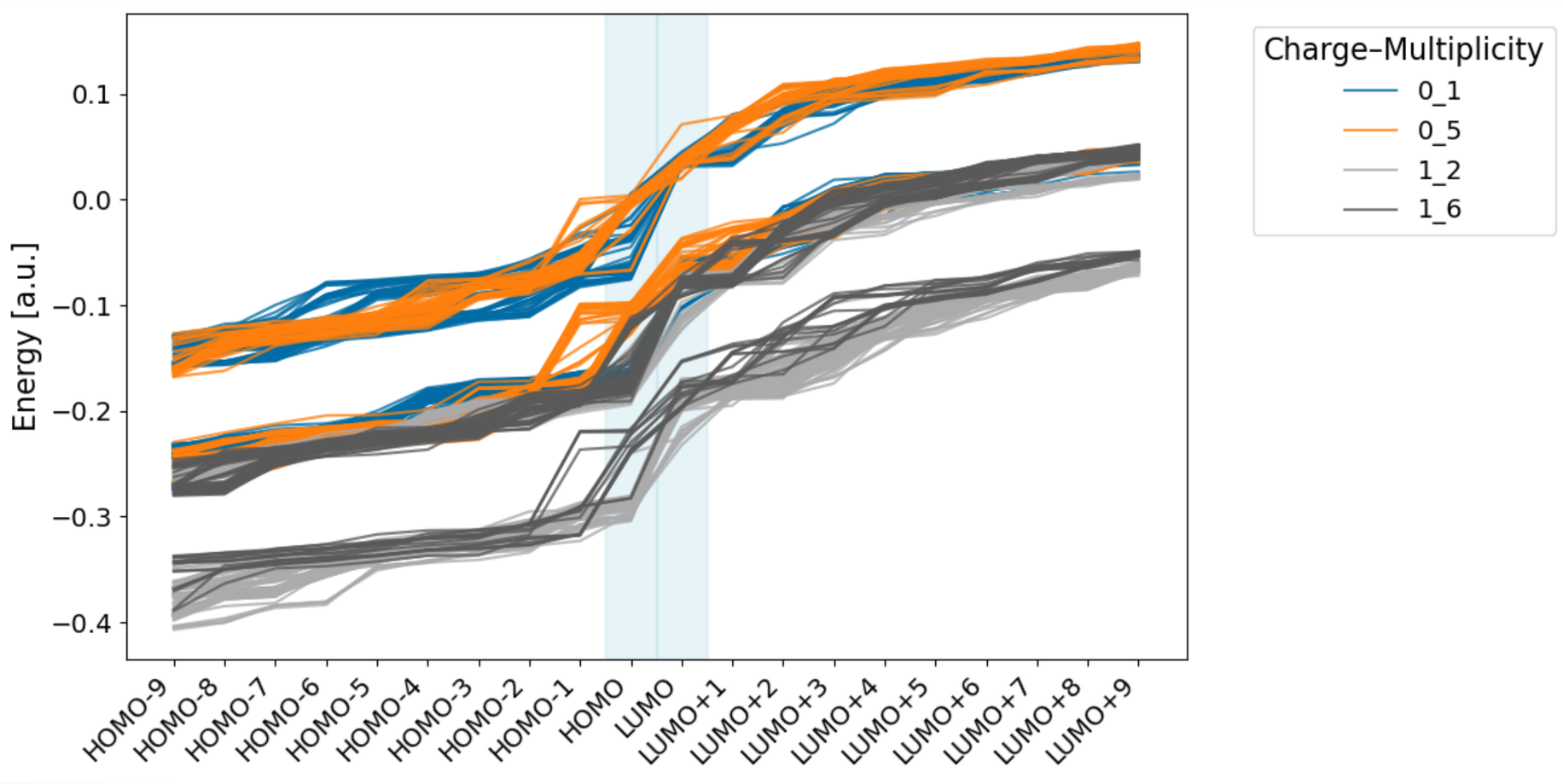
xTB: Bannwarth et al., Comp Mol Sci 2021

The Heme Electronic Structure Dataset HESD

Structural Diversity

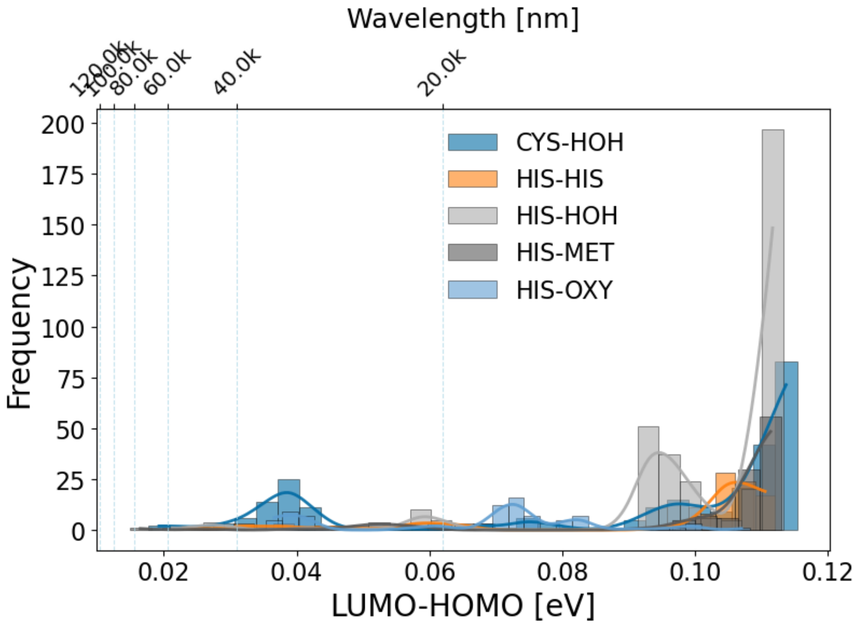


The Heme Electronic Structure Database

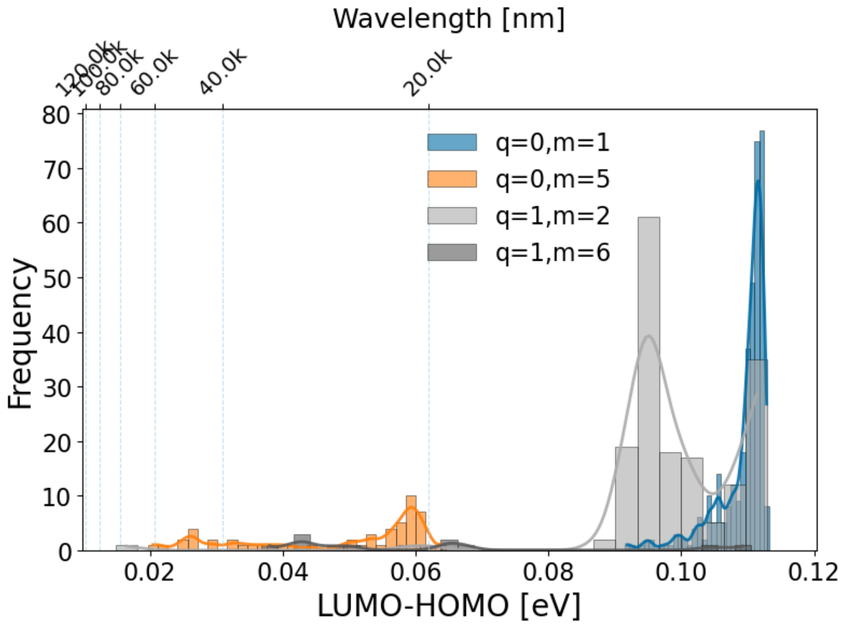


HOMO(-x)/LUMO(+x) orbital energies

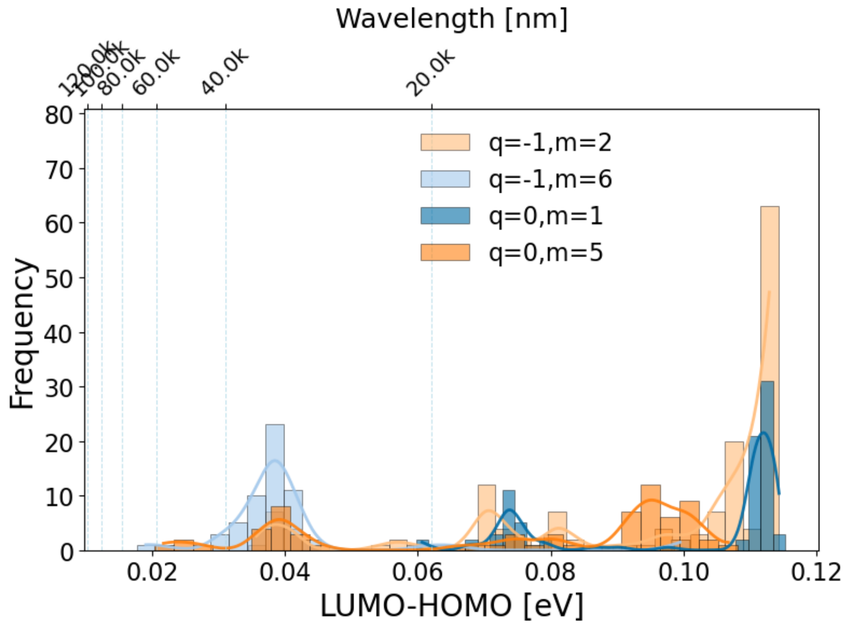
The Heme Electronic Structure Database



HOMO/LUMO orbital energy differences

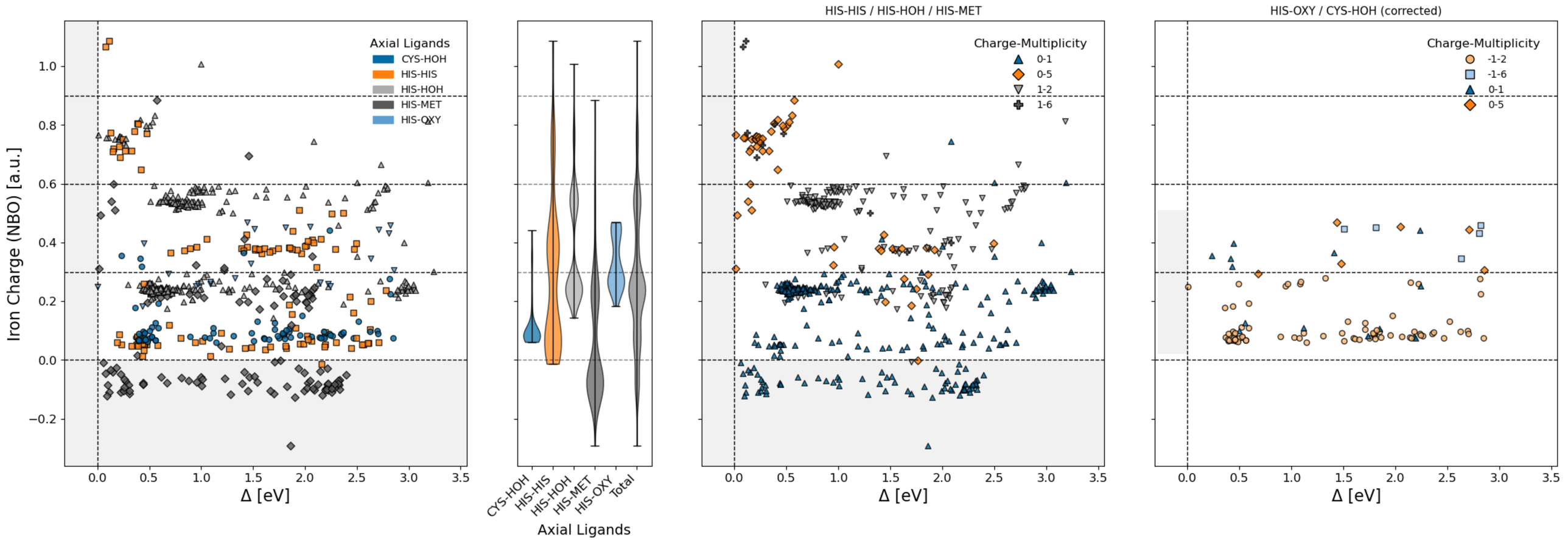


Only HIS-HIS/HIS-HOH/HIS-MET



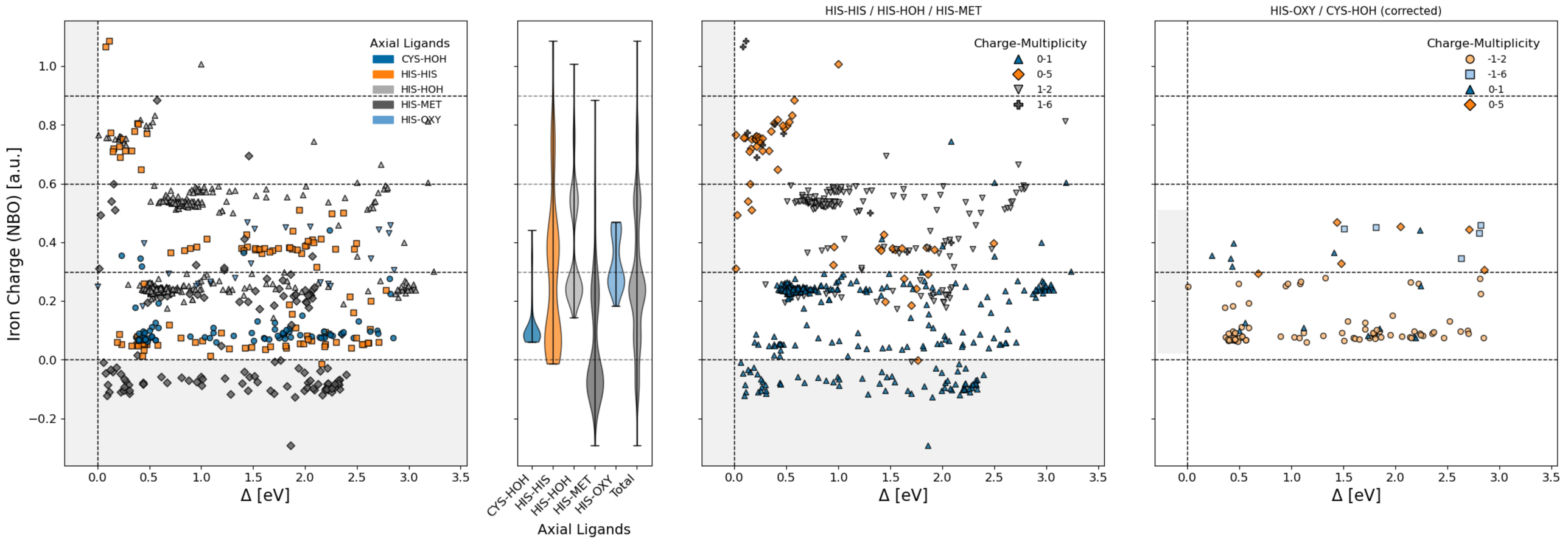
Only CYS-HOH/HIS-OXY

The Heme Electronic Structure Database



LFT Energy Splitting Δ - Iron Charge

The Heme Electronic Structure Database

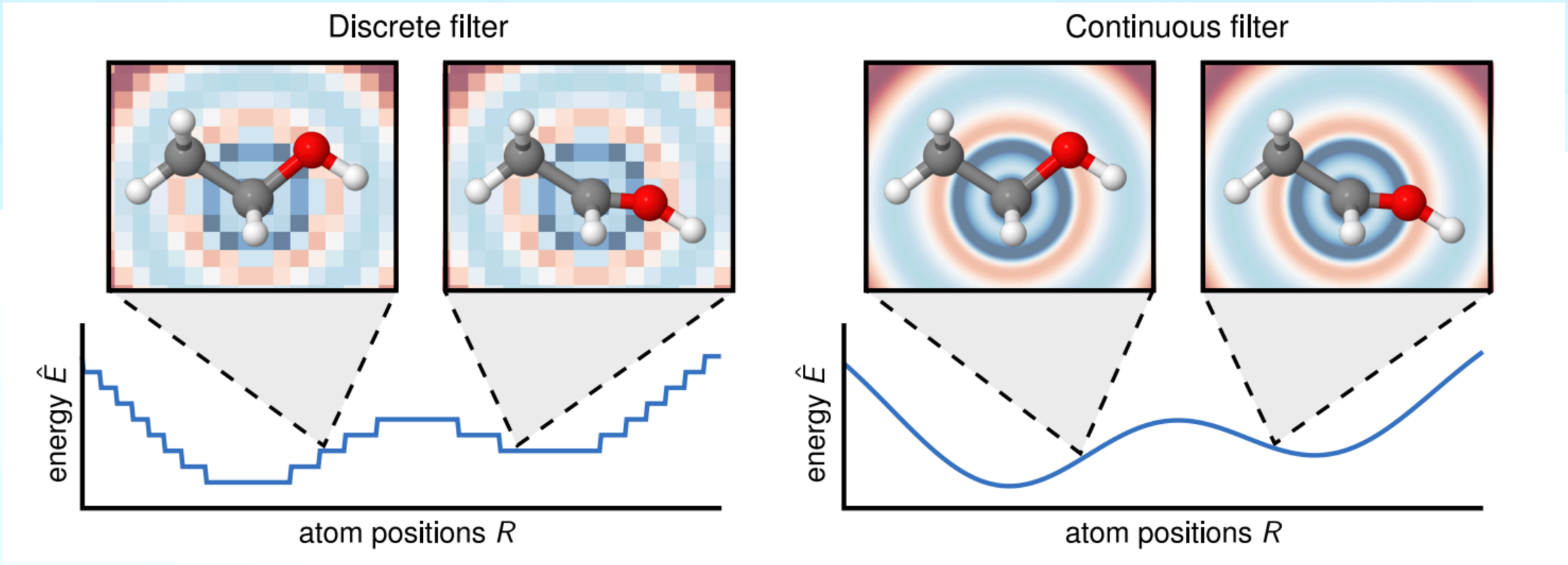


LFT Energy Splitting Δ - Iron Charge

There are correlations between orbital energies and charge distributions, but they are complex

SchNet

Schütt et al., 2018



Continuous Filter Convolutions

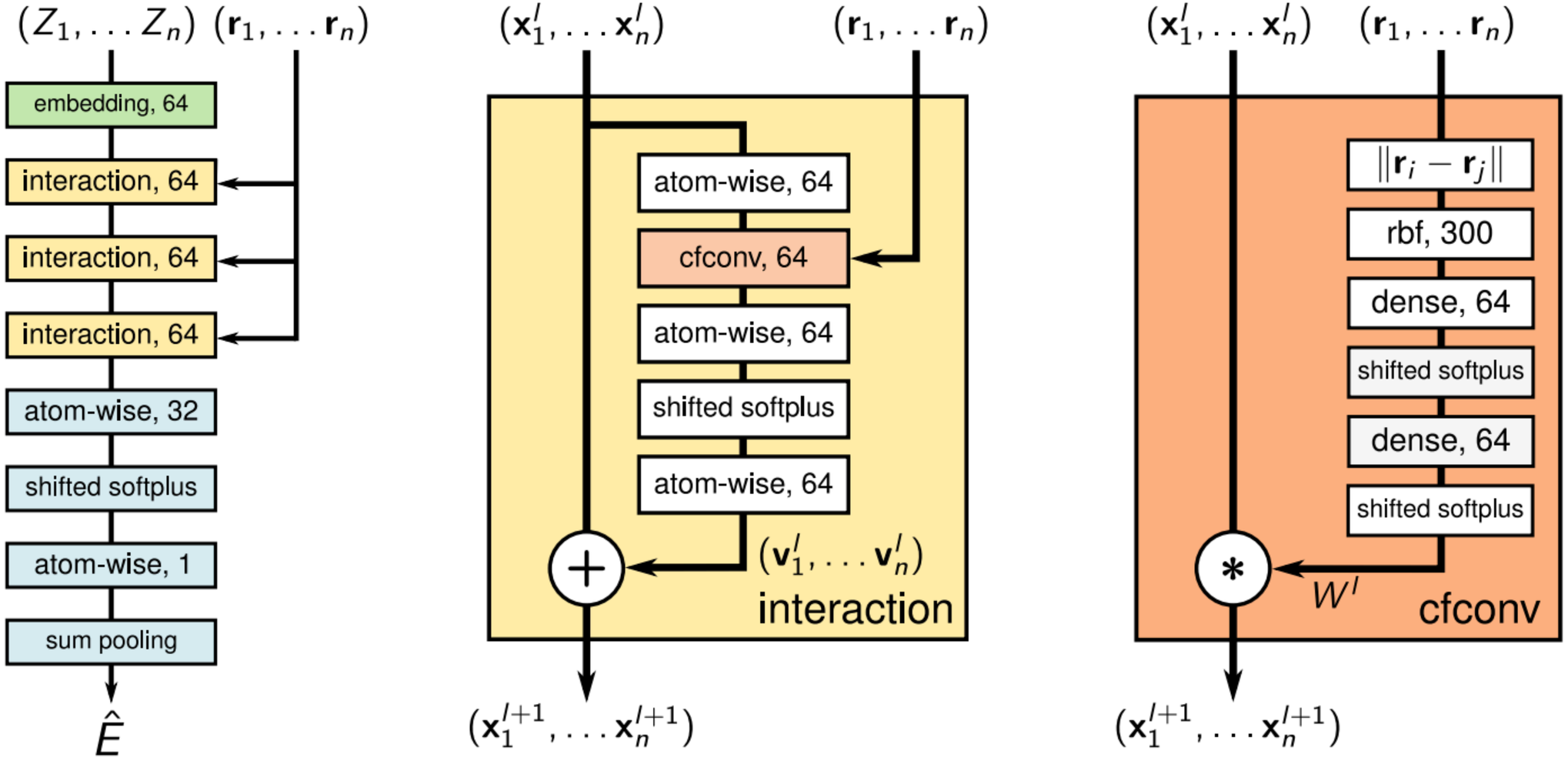
- Equal distance can be achieved in problems with unequal distance through resampling schemes, but only for some problems and with complete loss of some spacing information, or only with super large grid
- We require a filter generating function enabling continuous data by mapping from position to corresponding filter values

$$W^l : \mathbb{R}^D \rightarrow \mathbb{R}^F$$

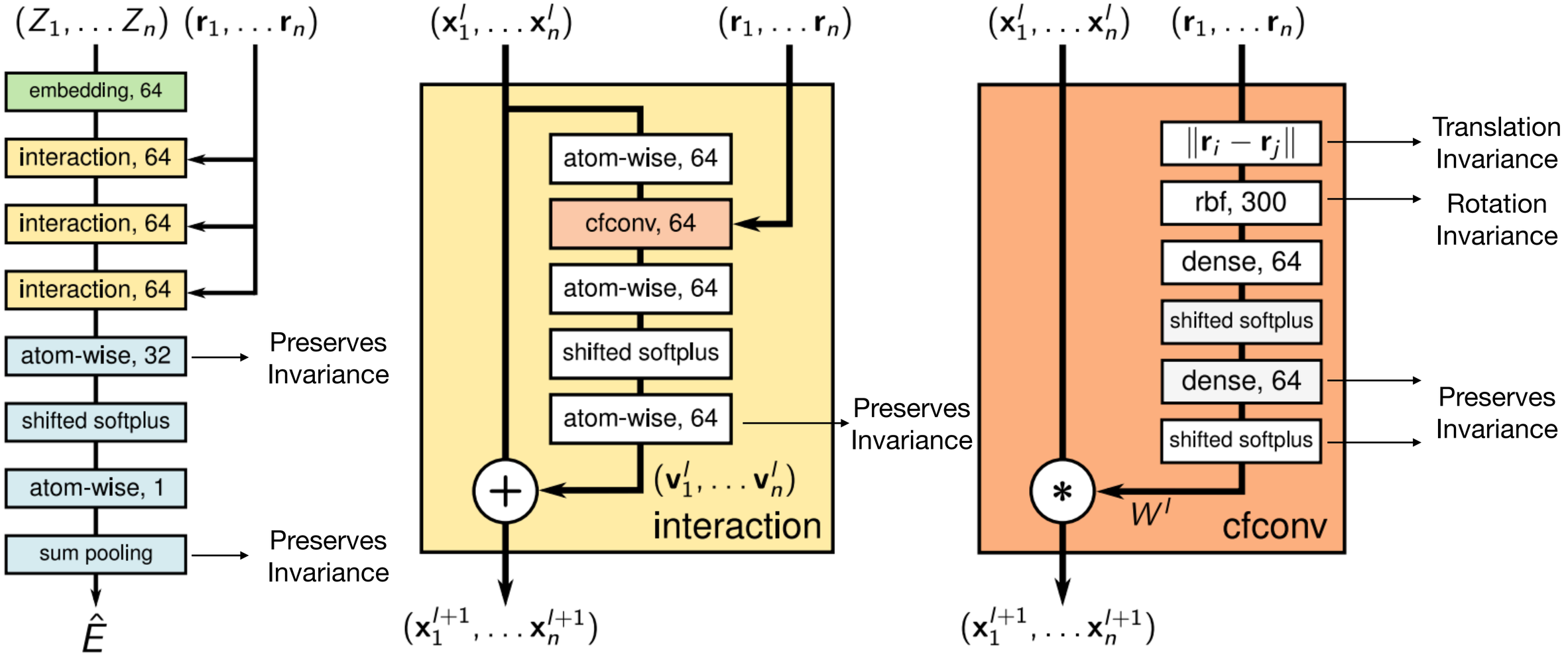
- Filter generating function is modelled through a neural network
- Input: atomic distances scaled with radial basis functions

$$x_i^{l+1} = (X^l * W^l) = \sum_j x_j^l \circ W^l(\mathbf{r}_i - \mathbf{r}_j)$$

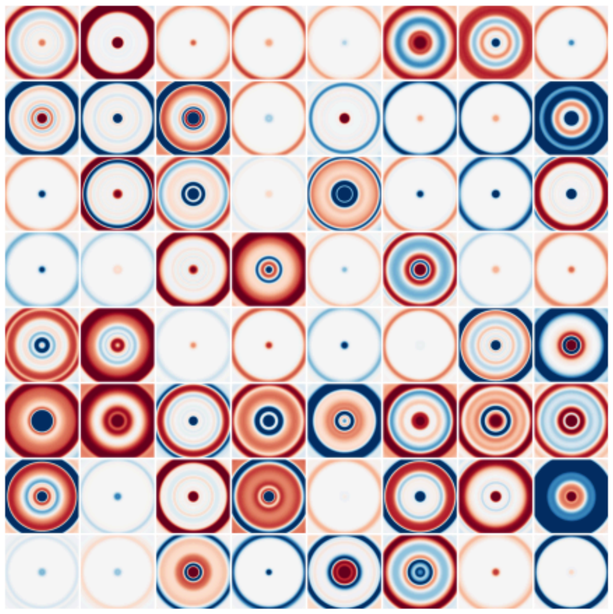
SchNet


$(Z_1, \ldots Z_n)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
embedding, 64
interaction, 64
interaction, 64
interaction, 64
atom-wise, 32
shifted softplus
atom-wise, 1
sum pooling
$\hat{E}$
$(\mathbf{x}_1^l, \ldots \mathbf{x}_n^l)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
atom-wise, 64
cfconv, 64
atom-wise, 64
shifted softplus
atom-wise, 64
$(\mathbf{v}_1^l, \ldots \mathbf{v}_n^l)$
interaction
$(\mathbf{x}_1^{l+1}, \ldots \mathbf{x}_n^{l+1})$
$(\mathbf{x}_1^l, \ldots \mathbf{x}_n^l)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
$\|\mathbf{r}_i - \mathbf{r}_j\|$
rbf, 300
dense, 64
shifted softplus
dense, 64
shifted softplus
W^l
cfconv
$(\mathbf{x}_1^{l+1}, \ldots \mathbf{x}_n^{l+1})$

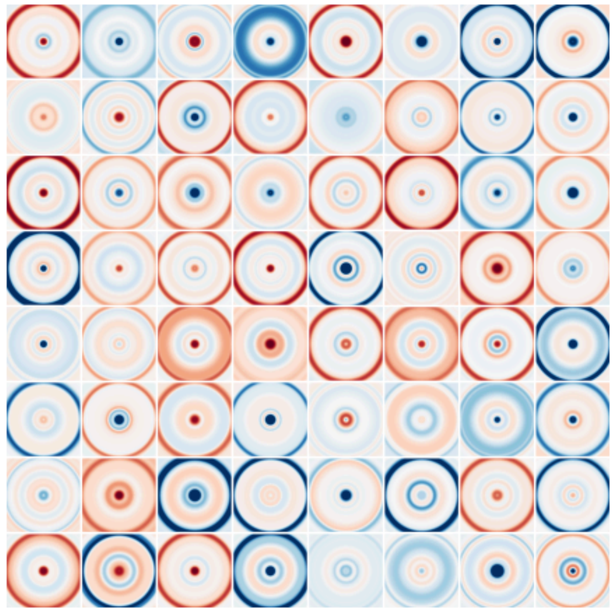
SchNet


$(Z_1, \ldots Z_n)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
embedding, 64
interaction, 64
interaction, 64
interaction, 64
atom-wise, 32
Preserves Invariance
shifted softplus
atom-wise, 1
sum pooling
Preserves Invariance
$\hat{E}$
$(\mathbf{x}_1^l, \ldots \mathbf{x}_n^l)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
atom-wise, 64
cfconv, 64
atom-wise, 64
shifted softplus
atom-wise, 64
$(\mathbf{v}_1^l, \ldots \mathbf{v}_n^l)$
interaction
Preserves Invariance
$(\mathbf{x}_1^{l+1}, \ldots \mathbf{x}_n^{l+1})$
$(\mathbf{x}_1^l, \ldots \mathbf{x}_n^l)$
$(\mathbf{r}_1, \ldots \mathbf{r}_n)$
$\|\mathbf{r}_i - \mathbf{r}_j\|$
rbf, 300
dense, 64
shifted softplus
dense, 64
shifted softplus
W^l
cfconv
Translation Invariance
Rotation Invariance
Preserves Invariance
$(\mathbf{x}_1^{l+1}, \ldots \mathbf{x}_n^{l+1})$

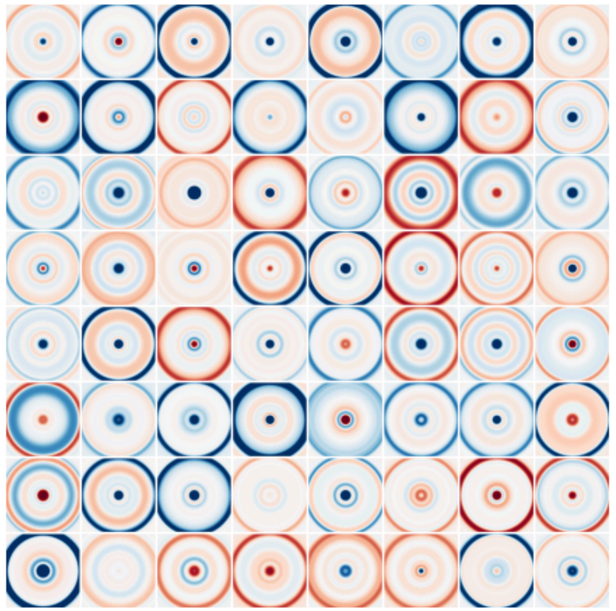
Continuous Representations



(a) 1st interaction block



(b) 2nd interaction block



(c) 3rd interaction block

SchNet Learning

- Obtain an energy conserving force model by differentiating the energy model w.r.t atom positions

$$\hat{\mathbf{F}}_i(Z_1, \ldots, Z_n, \mathbf{r}_i, \ldots, \mathbf{r}_n) = \frac{\partial \hat{E}}{\partial \mathbf{r}_i}(Z_1, \ldots, Z_n, \mathbf{r}_i, \ldots, \mathbf{r}_n)$$

- And loss function including total energy E and forces $\mathbf{F}_i$ to perform well in both properties

$$l(\hat{E}, (E, \mathbf{F}_1, \ldots, \mathbf{F}_n)) = \rho \|E - \hat{E}\|^2 + \frac{1}{n} \sum_{i=0}^{n} \|\mathbf{F}_i - \left(-\frac{\partial \hat{E}}{\partial \hat{R}_i}\right)\|^2$$

SchNet Performance

QM9 dataset (small organic molecules, chemical d.o.f.)

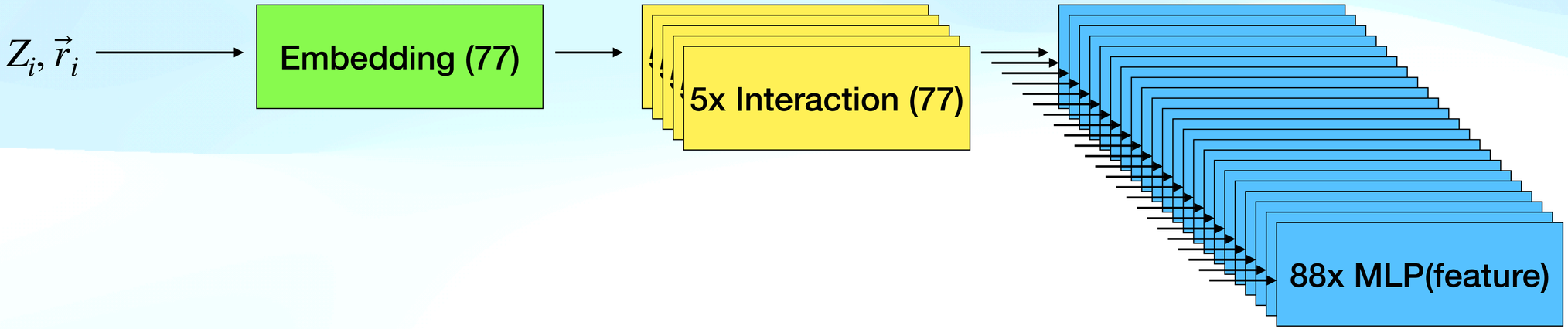
N	SchNet	DTNN [20]	enn-s2s [21]	enn-s2s-ens5 [21]
50,000	**0.59**	0.94	–	–
100,000	**0.34**	0.84	–	–
110,462	**0.31**	–	0.45	0.33

ISO17 (conformational & chemical changes, MD of 129 isomers present in QM9)

		mean predictor	**SchNet**	
			energy	*energy+forces*
known molecules / unknown conformation	*energy*	14.89	0.52	**0.36**
	forces	19.56	4.13	**1.00**
unknown molecules / unknown conformation	*energy*	15.54	3.11	**2.40**
	forces	19.15	5.71	**2.18**

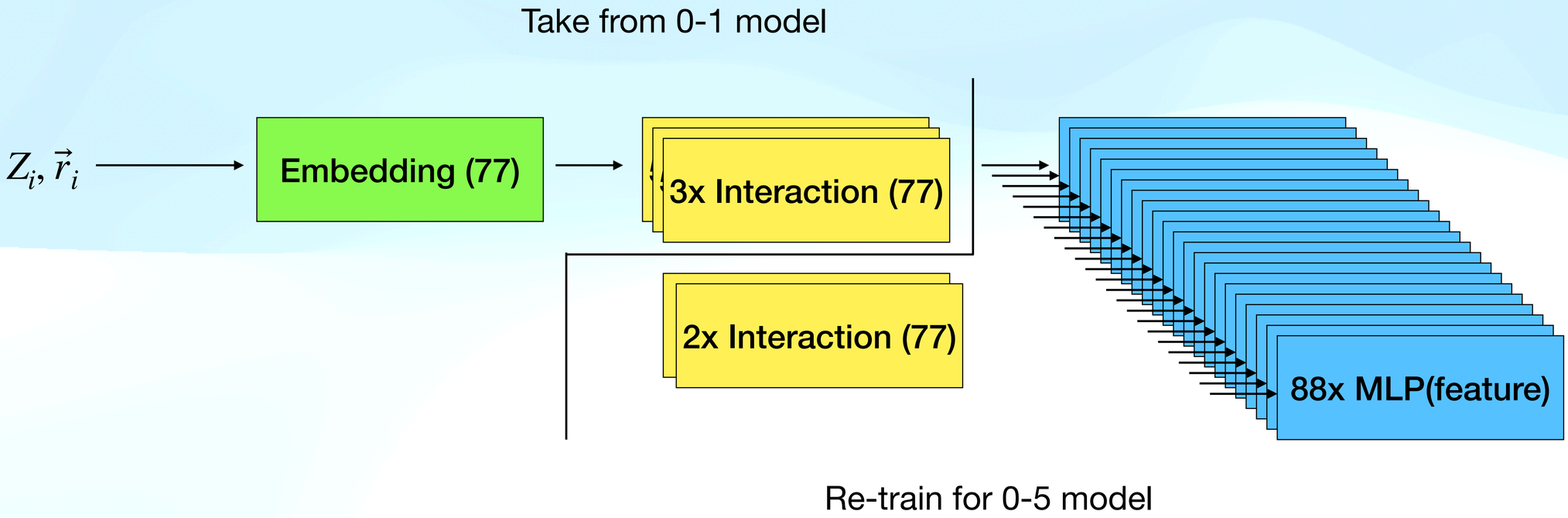
HemeSchNet

Architecture Base Model (charge=0, multiplicity=1)



HemeSchNet

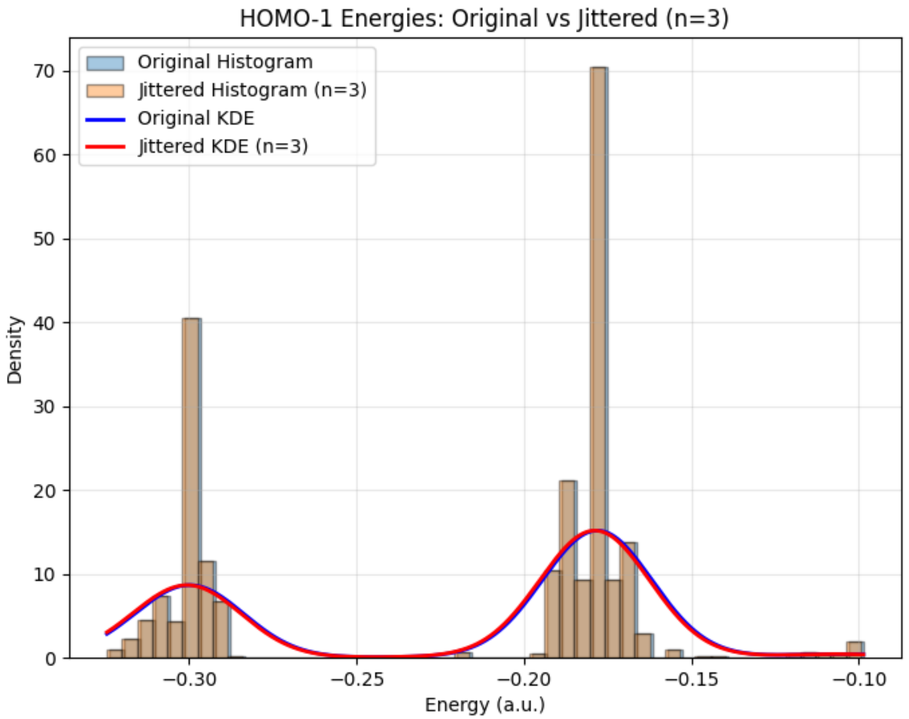
Architecture derived Models (e.g. charge=0, multiplicity=5)



HemeSchNet

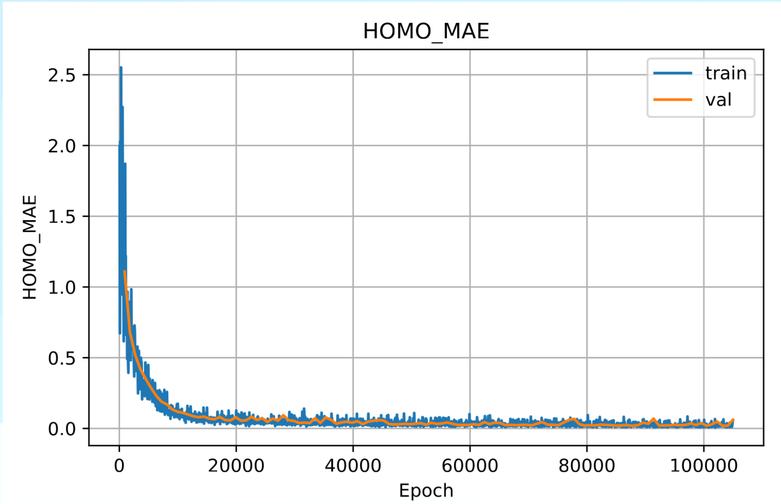
- Machine Learning Problems in Chemistry are usually constrained by data availability
- Experiments are insanely costly
- Simulations are costly
- Only ML is cheap
- The dataset is sparse, hence we oversampled around the points in the latent space we know about

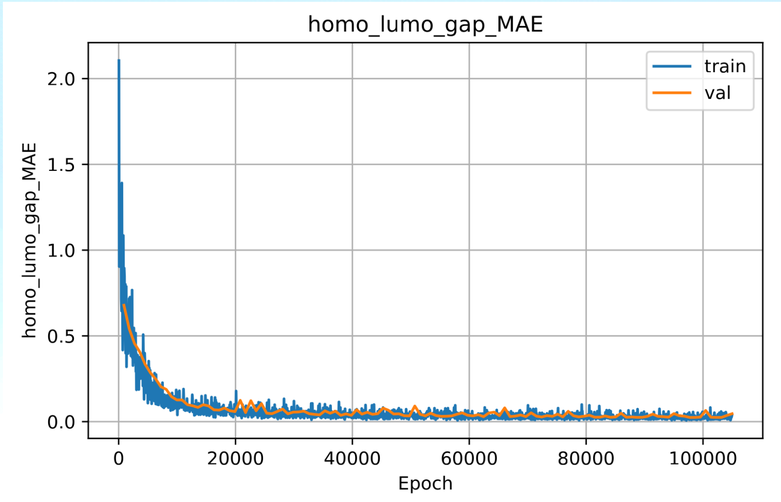
$$x_{\vec{r},jittered} = x_{\vec{r}} + \varepsilon, \varepsilon \sim \mathcal{N}(0,0.02)$$

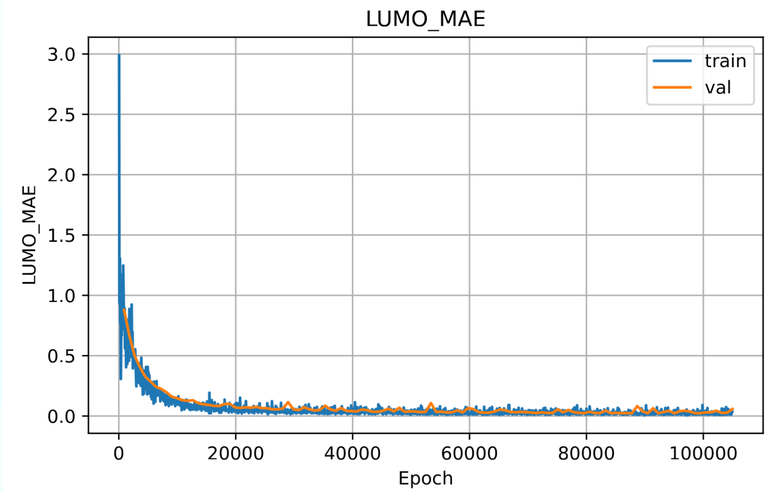


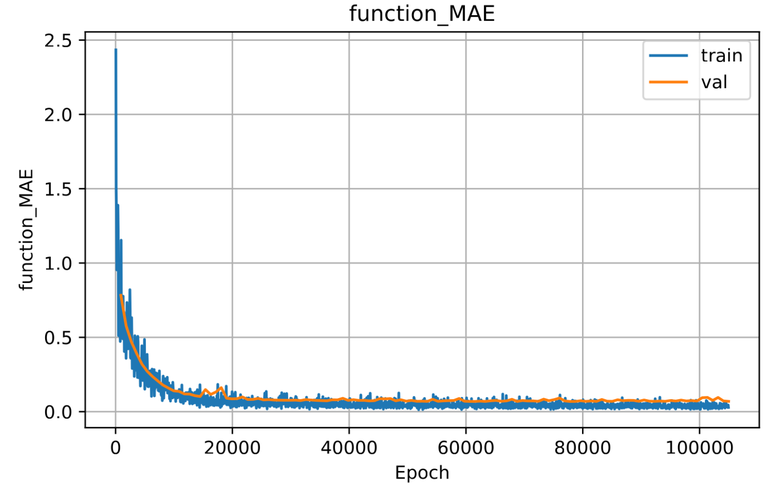
HemeSchNet

Jones et al., in preparation

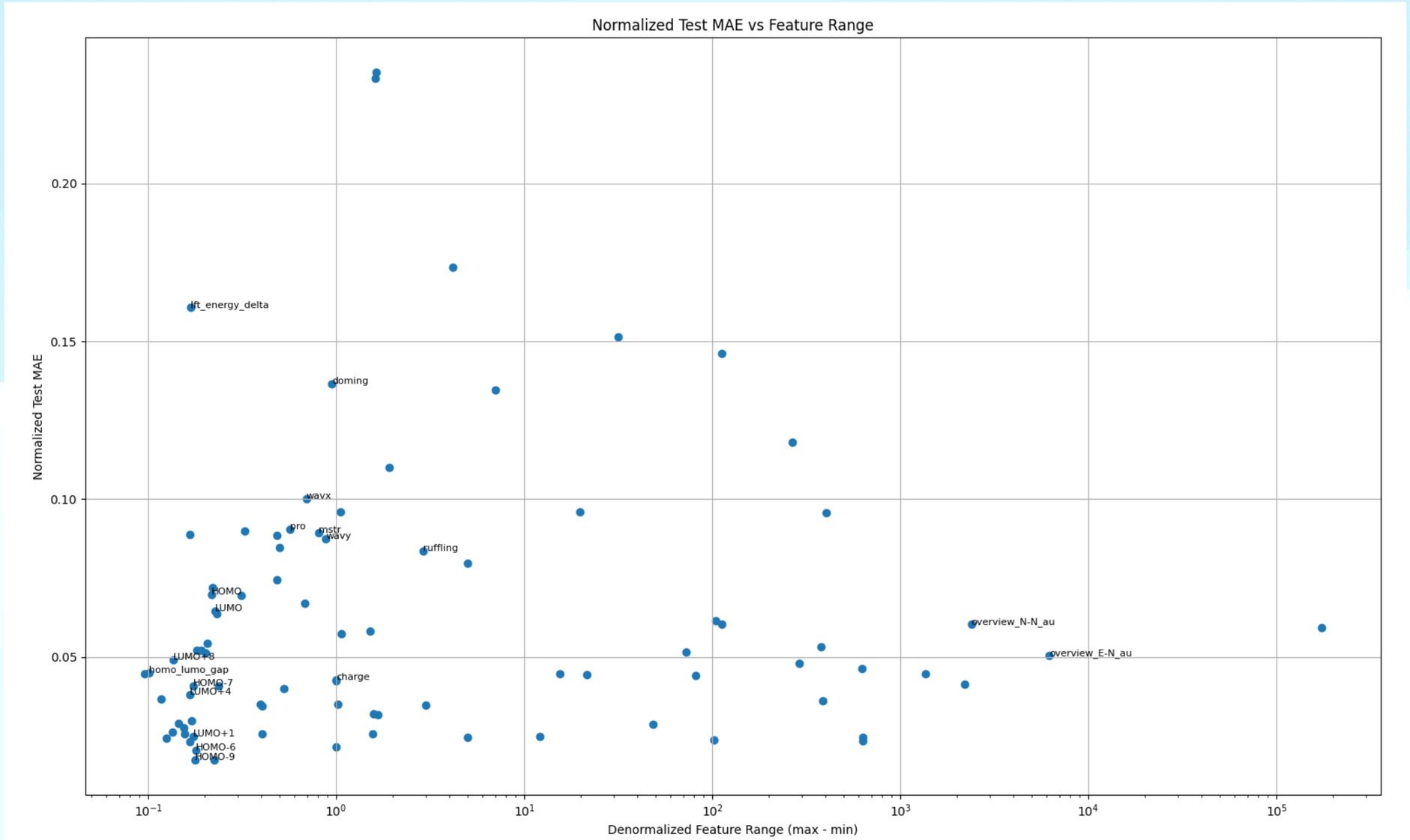








HemeSchNet


Normalized Test MAE vs Feature Range
Normalized Test MAE
Denormalized Feature Range (max - min)
0.20
0.15
0.10
0.05
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
lft_energy_delta
doming
wavx
pro
mstr
wavy
ruffling
HOMO
LUMO
overview_N-N_au
overview_E-N_au
LUMO+8
homo_lumo_gap
charge
HOMO-7
LUMO+4
LUMO+1
HOMO-6
HOMO-9

Acknowledgements

Contact: j.jones@tu-berlin.de

unisyscat

Technische Universität Berlin

SFB 1078

Mroginski Group: Biomodeling @ TU Berlin
https://github.com/biomodeling-tub