

Beyond Jacob's Ladder: A Physics-driven AI Approach to Density Functional Theory

Peter Halat^{1*}, Florencia Stella¹, Russhil Khurana¹, Sophie A. Crouch,¹ Nicole Wang,¹ Ekaterina I.
Izgorodina^{1*}

¹ School of Chemistry, Monash University, Clayton, Victoria, Australia.

*Corresponding authors. E-mails: peter.halat@monash.edu, katya.pas@monash.edu

Supplementary Information

Table S1: Performance of the ML model after being trained on atomisation energies (AEs) and interaction energies (EINTs) across 65% training, 17.5% validation and 17.5% test data subsets. All values are given in kcal/mol.

	Training			Validation			Test		
	AEs	EINT	Both	AEs	EINT	Both	AEs	EINT	Both
MAE	0.46	0.5	0.47	0.5	0.51	0.5	0.51	0.47	0.5
RMSE	0.72	0.8	0.73	0.87	0.7	0.84	1.03	0.6	0.98
Maximum	10.43	9.47	10.43	12.56	2.88	12.56	14.49	2.11	14.49

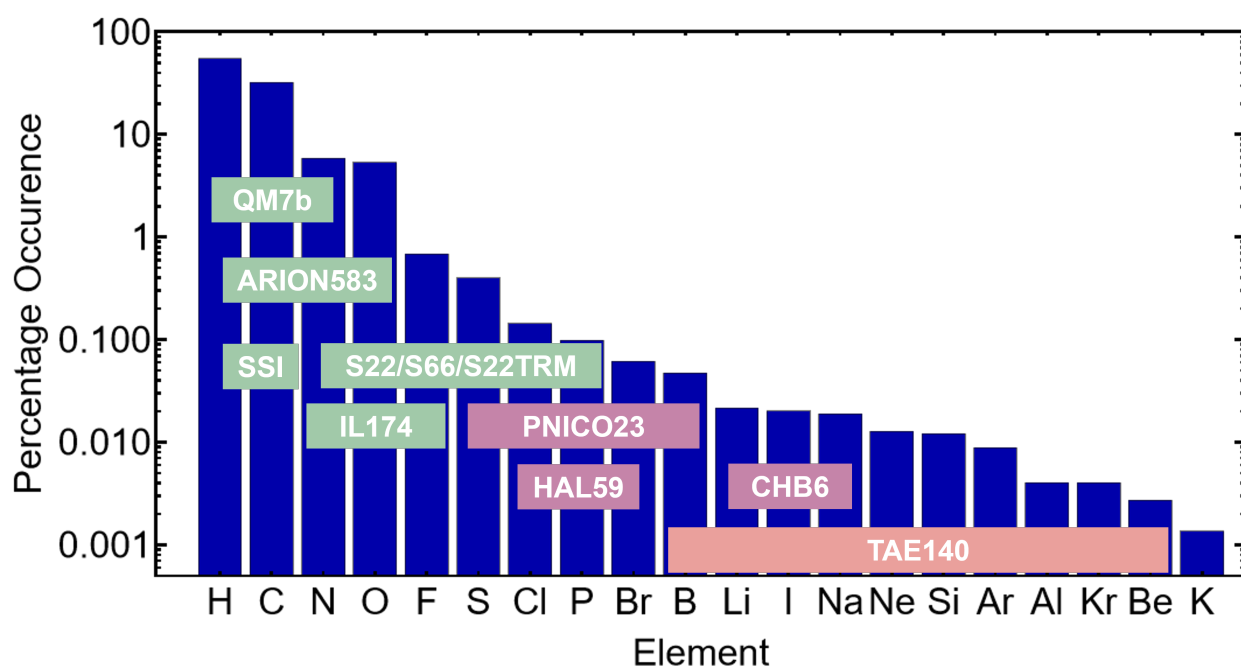


Figure S1: Distribution of chemical elements present across all chemical systems constituting the atomisation and interaction energy datasets used in this study.

Table S2: (Left) Performance of the ML model after being trained on atomisation energies (AEs) and interaction energies (EINTs) on reaction energy (REs) and reaction barrier (RBs) datasets. (Right) Performance of the model after being trained on AEs, EINTs, REs and RBs on reaction energy (REs) and reaction barrier (RBs) datasets. All values are given in kcal/mol.

	As unseen data			After Training					
	REs	RBs		Training		Validation		Test	
	REs	RBs		REs	RBs	REs	RBs	REs	RBs
MAE	3.88	4.07	MAE	1.47	1.32	1.55	1.49	1.47	1.94
RMSE	5.70	5.14	RMSE	1.92	1.85	2.11	1.94	1.94	2.43
Maximum	25.43	13.09	Maximum	6.56	5.91	7.6	5.24	5.39	6.26

Table S3: Retained performance of the model on atomisation energies (AEs) and interaction energies (EINTs) after being additionally trained on reaction energies and reaction barriers. All values are given in kcal/mol.

	Training			Validation			Test		
	AEs	EINT	Both	AEs	EINT	Both	AEs	EINT	Both
MAE	0.54	0.54	0.54	0.58	0.53	0.57	0.58	0.52	0.57
RMSE	0.75	0.82	0.76	1.00	0.71	0.96	1.15	0.67	1.09
Maximum	8.51	8.22	8.51	20.58	2.56	20.58	22.78	2.46	22.78

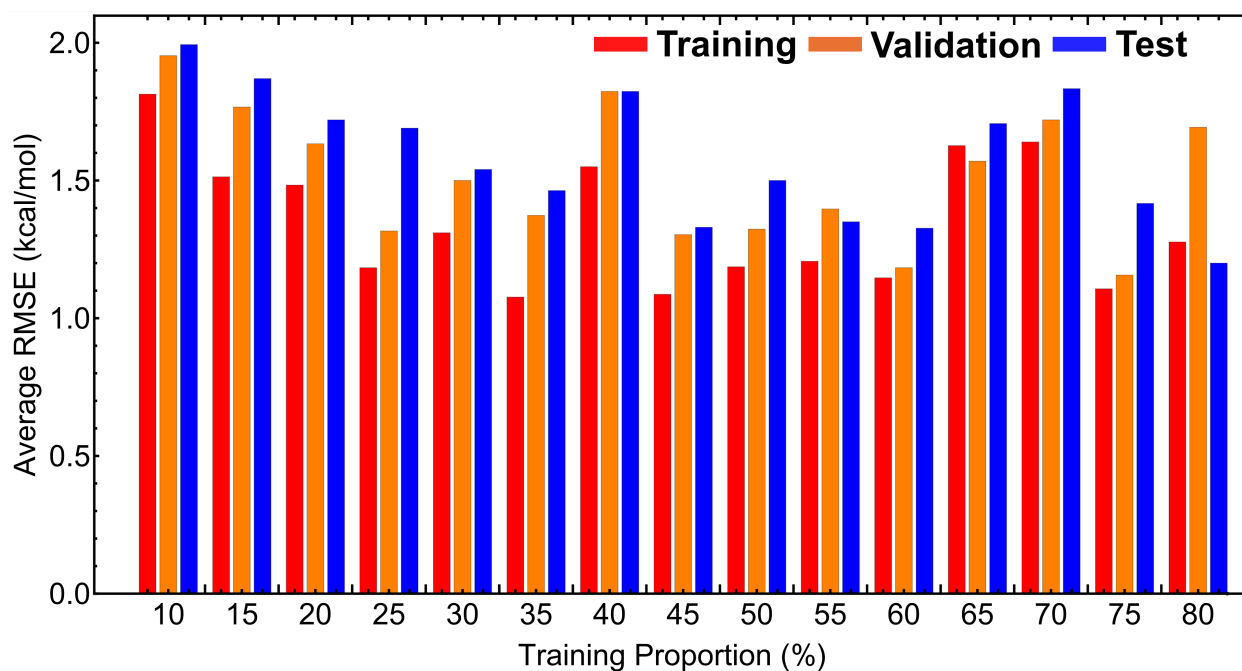


Figure S2: Average RMSE of training (red), validation (orange) and test (blue) subsets across a training proportion sweep between 10% to 80% of atomisation energies and interaction energies. Averages were obtained from triplicate runs.