Electronic Supplementary Material (ESI) for Digital Discovery.

Supporting Information for *High-throughput Quantum Theory of Atoms in Molecules (QTAIM) for Geometric Deep Learning of Molecular and Reaction Properties*

Santiago Vargas,[†] Winston Gee,[†] and Anastassia Alexandrova[*,†,‡,¶]

† *Department of Chemistry and Biochemistry, University of California, Los Angeles, 90095-1569, USA*

‡*Department of Materials Science and Engineering, University of California, Los Angeles, 90095-1569, USA*

¶*California NanoSystems Institute*

E-mail: santiagovargas921@gmail.com,ana@chem.ucla.edu

Full set of QTAIM descriptors

Atom Descriptors	**Bond Descriptors**
Total Electrostatic Potential Φ_{tot}	Total Electrostatic Potential Φ_{tot}
Nuclear Electrostatic Potential Φ_{nuc}	Nuclear Electrostatic Potential Φ_{nuc}
Electronic Electrostatic Potential Φ_e	Electronic Electrostatic Potential Φ_e
Lagrangrian ($\nabla^2\rho$)	Lagrangrian ($\nabla^2\rho$)
Kinetic Energy Hamiltonian	Kinetic Energy Hamiltonian
Gradient Norm	Gradient Norm
Δ_g promolecular	Δ_g promolecular
Δ_g Hirshfield	Δ_g Hirshfield
Electron Density	Electron Density$)
Laplacian Electron Density	Laplacian Electron Density
Hessian Determinant	Hessian Determinant
Electron Localization Function(ELF)	Electron Localization Function(ELF)
Laplacian Norm	Laplacian Norm
Hessian eigenvalue (1st)	Hessian eigenvalue (1st)
Electronc Ellipticity	Electronc Ellipticity
Average Location Ion E	Average Location Ion E
Eta	Eta
Localized Orbital Locator	Localized Orbital Locator
energy density	energy density
α spin	α spin
β spin	β spin
spin density	spin density

Dataset Visualizations

Uncorrected Energies - LIBE

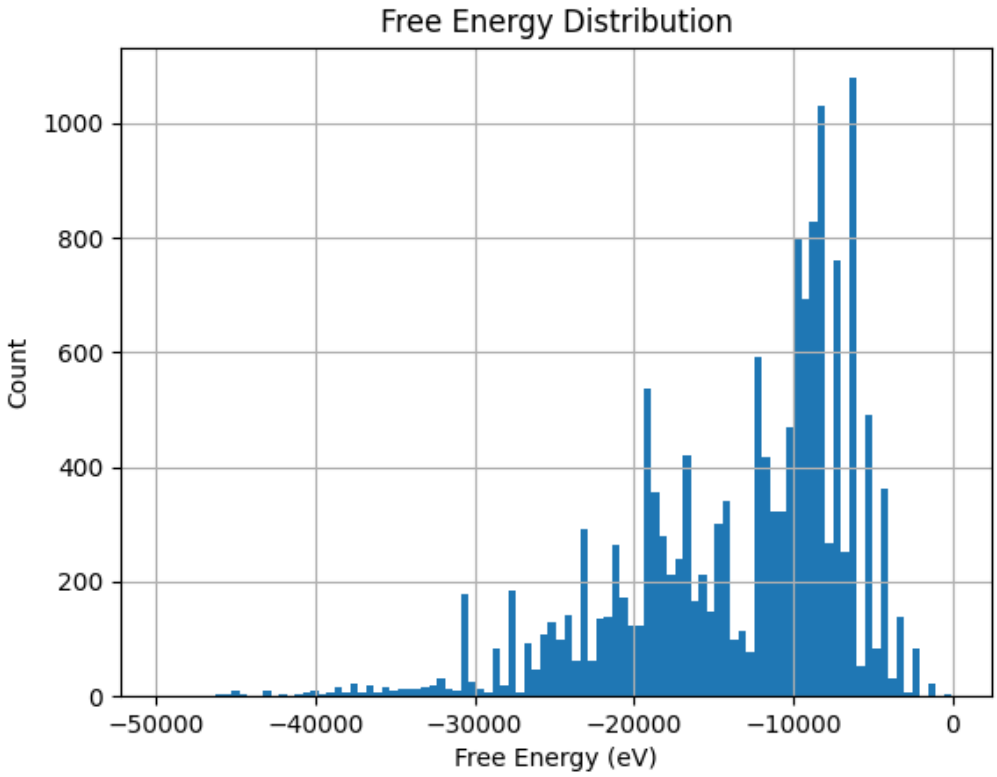

Figure S2: LIBE energies prior to correction

Corrected Energies - LIBE

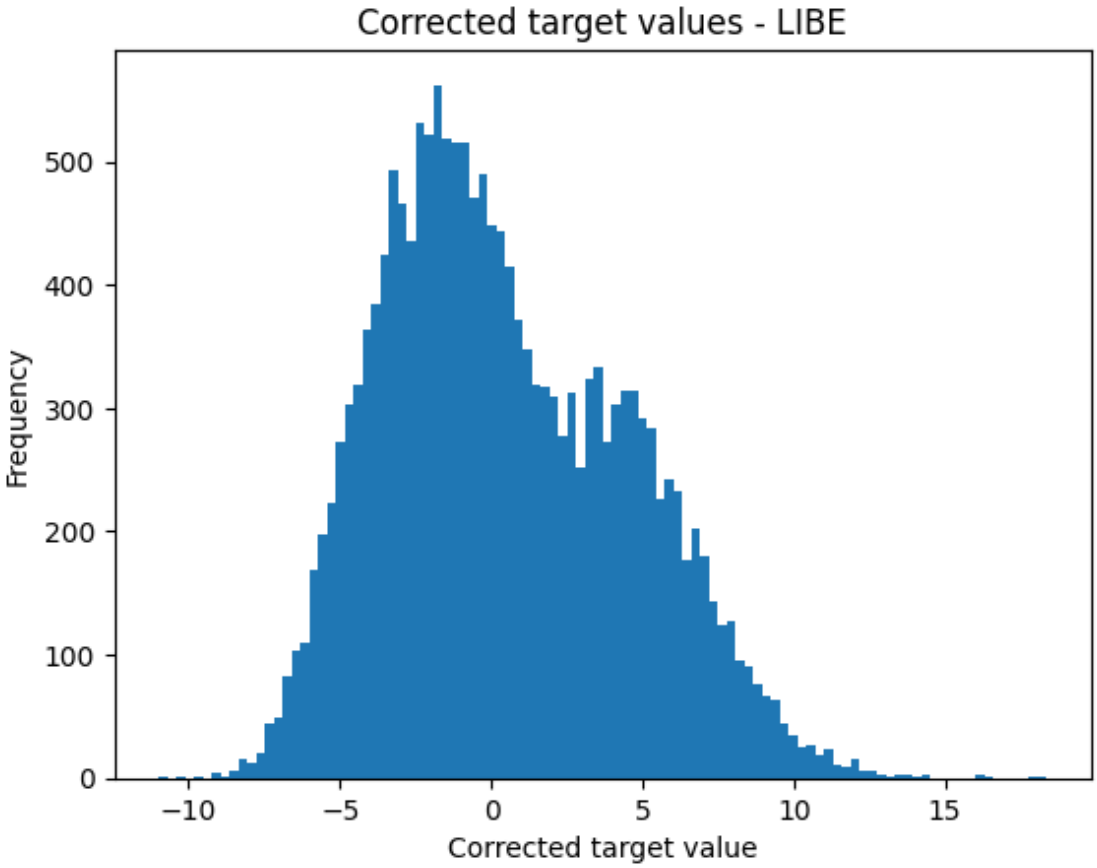

Figure S3: LIBE corrected energies

Correction Values

Table S4: Correction values used from raw LIBE energies

Atomic Number	Correction Value
1	-16.77537562
3	-206.45292515
6	-1034.69861041
7	-1488.80081496
8	-2048.19270236
9	-2717.83725543
15	-9286.36995521
16	-10831.57826394

Parity Plots

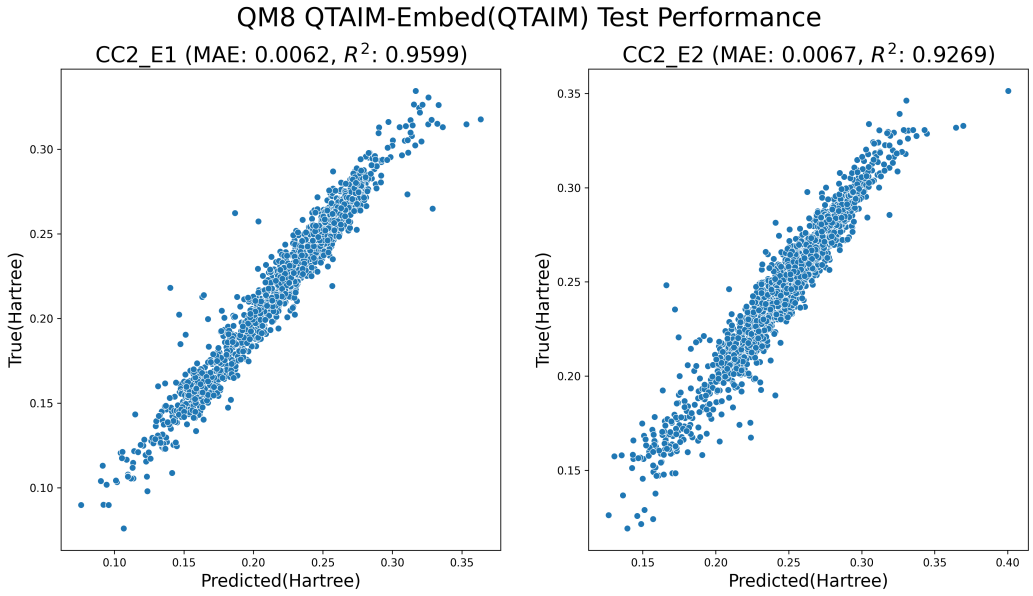

Figure S5: QM8 QTAIM test Partity.

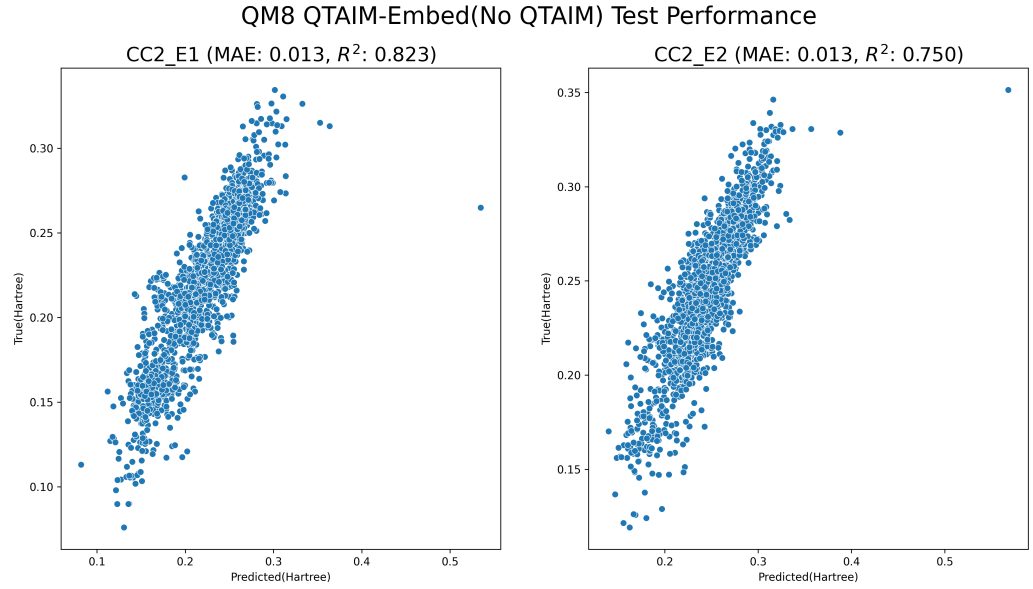

Figure S6: QM8 non-QTAIM test Partity.

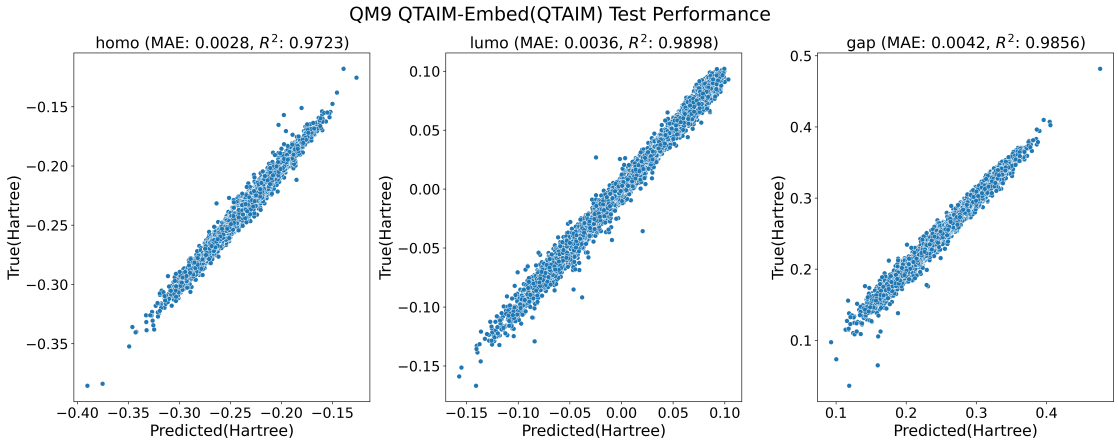

Figure S7: QM9 QTAIM test Partity.

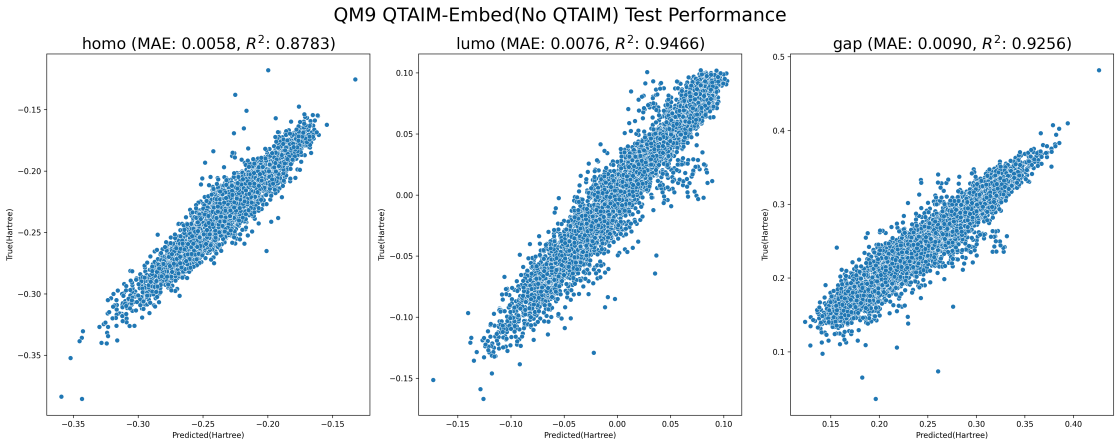

Figure S8: QM9 non-QTAIM test Partity.

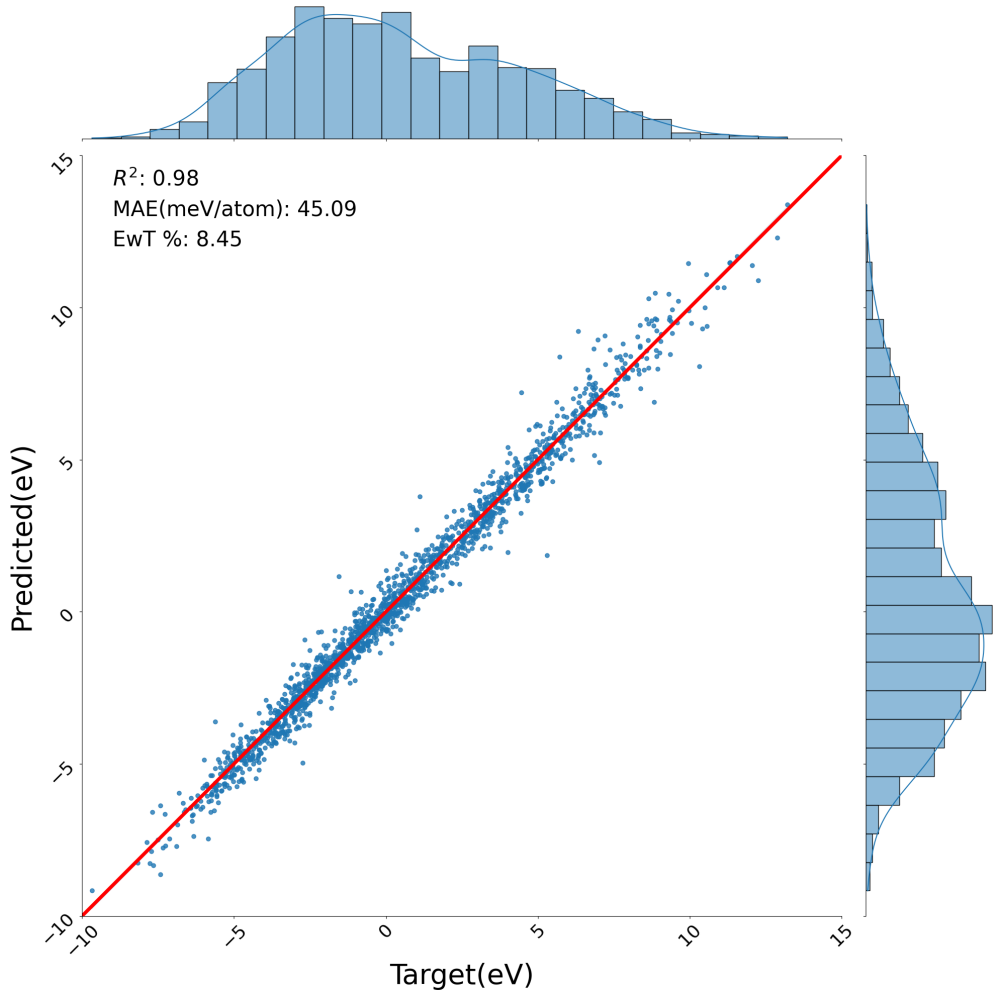

Figure S9: LIBE QTAIM test Partity.

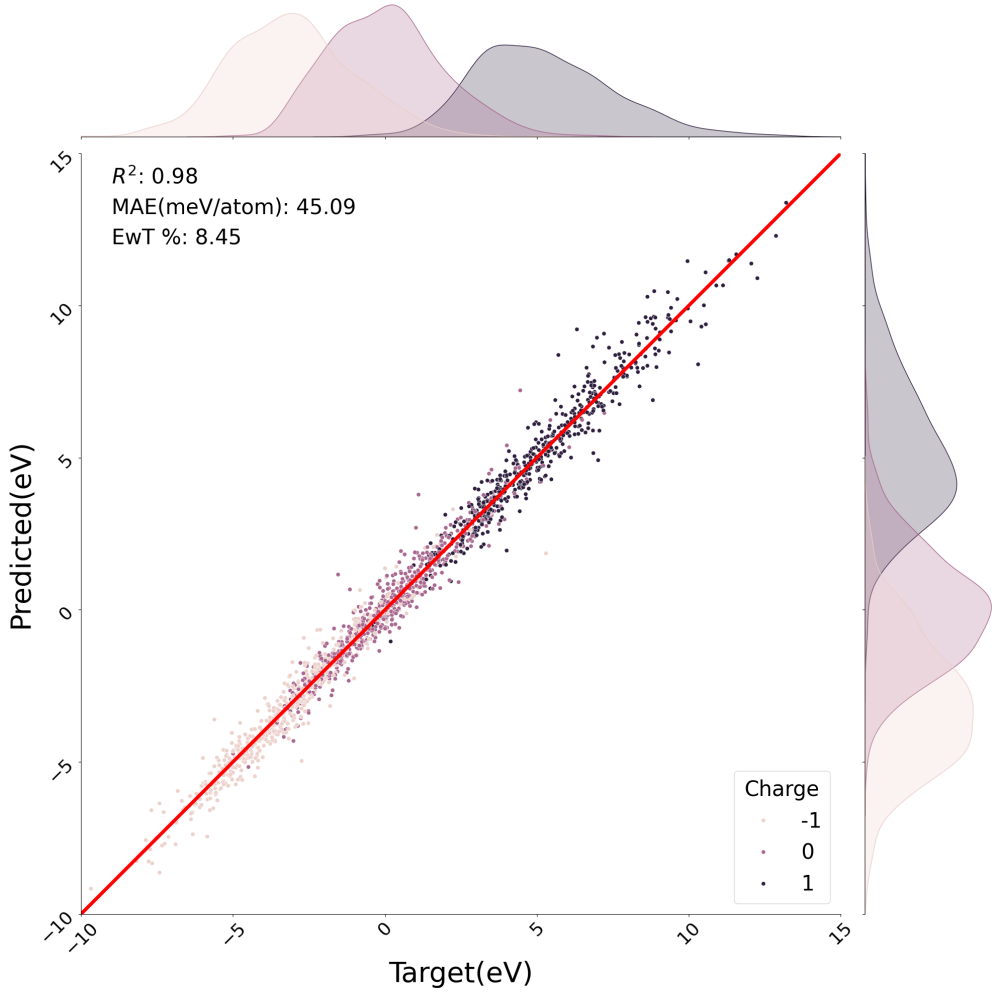

Figure S10: LIBE QTAIM test Partity, charge-partitioned.

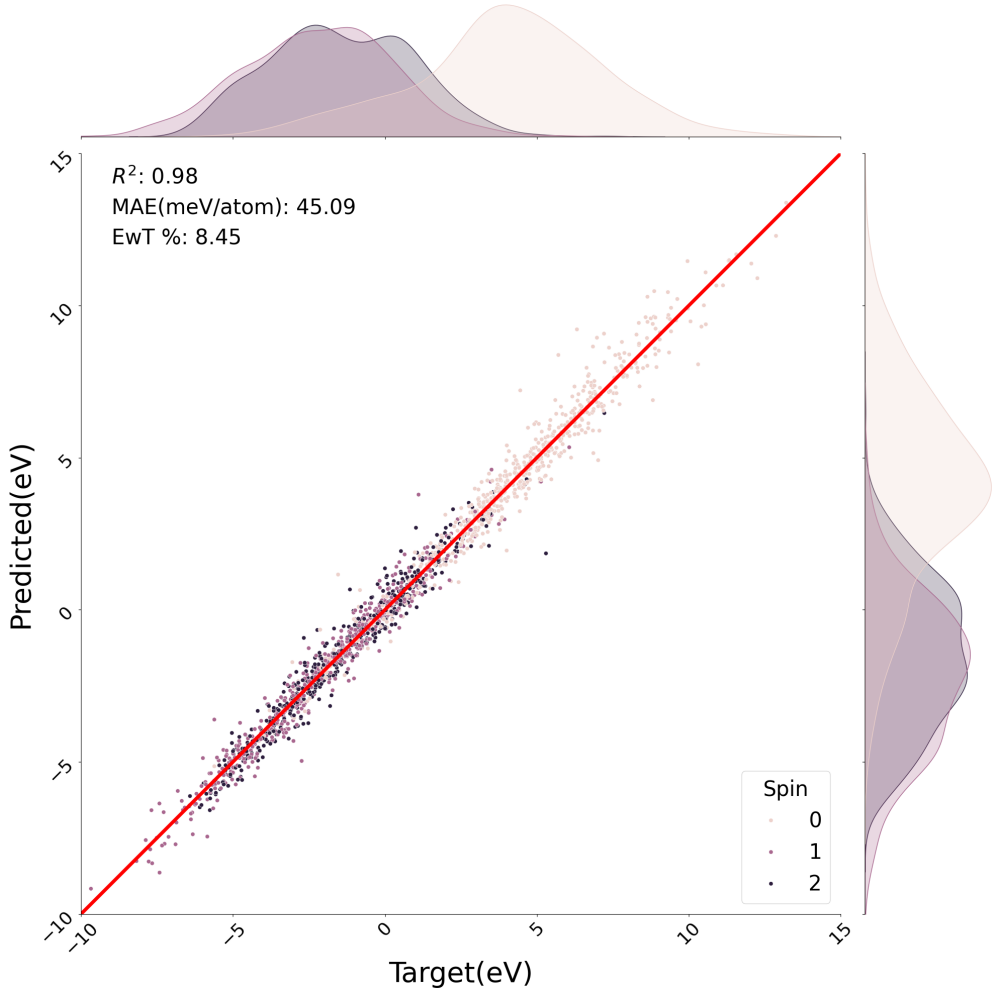

Figure S11: LIBE QTAIM test Partity, spin-partitioned.

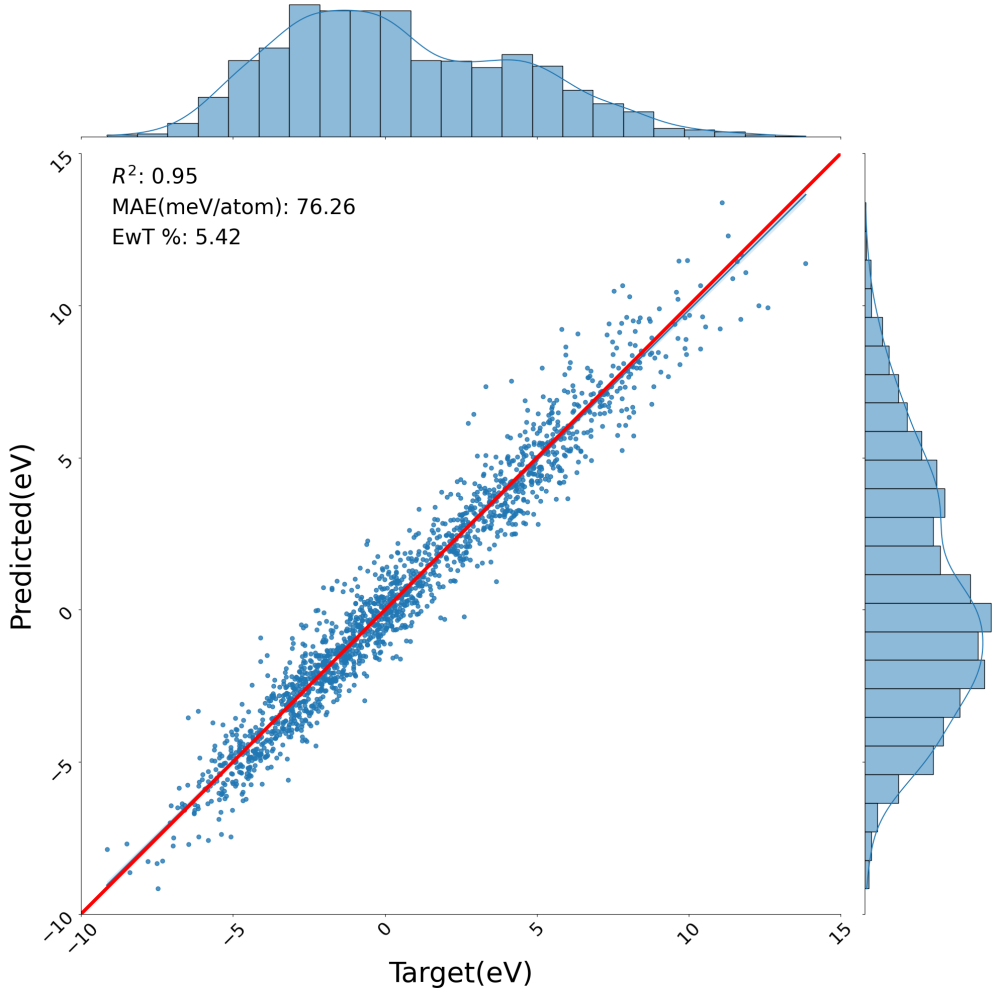

Figure S12: LIBE non-QTAIM test Partity.

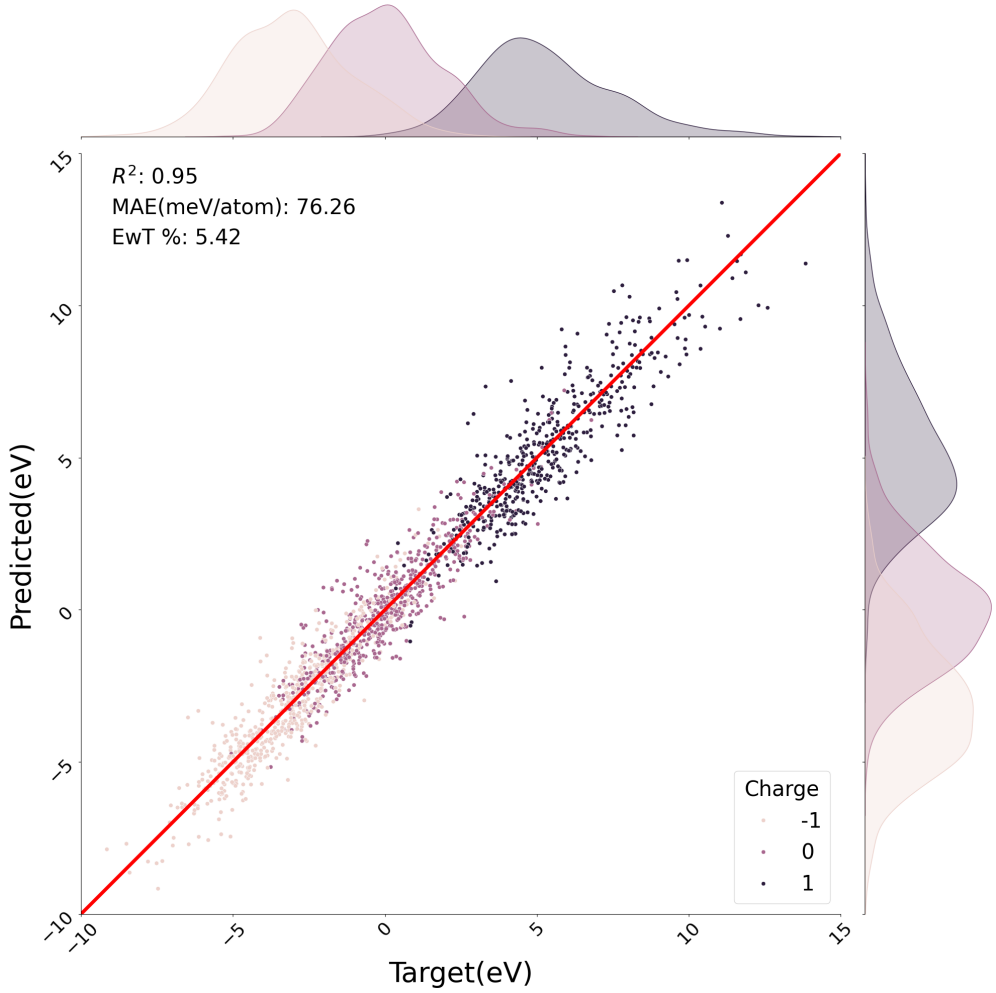

Figure S13: LIBE non-QTAIM test Partity, charge-partitioned.

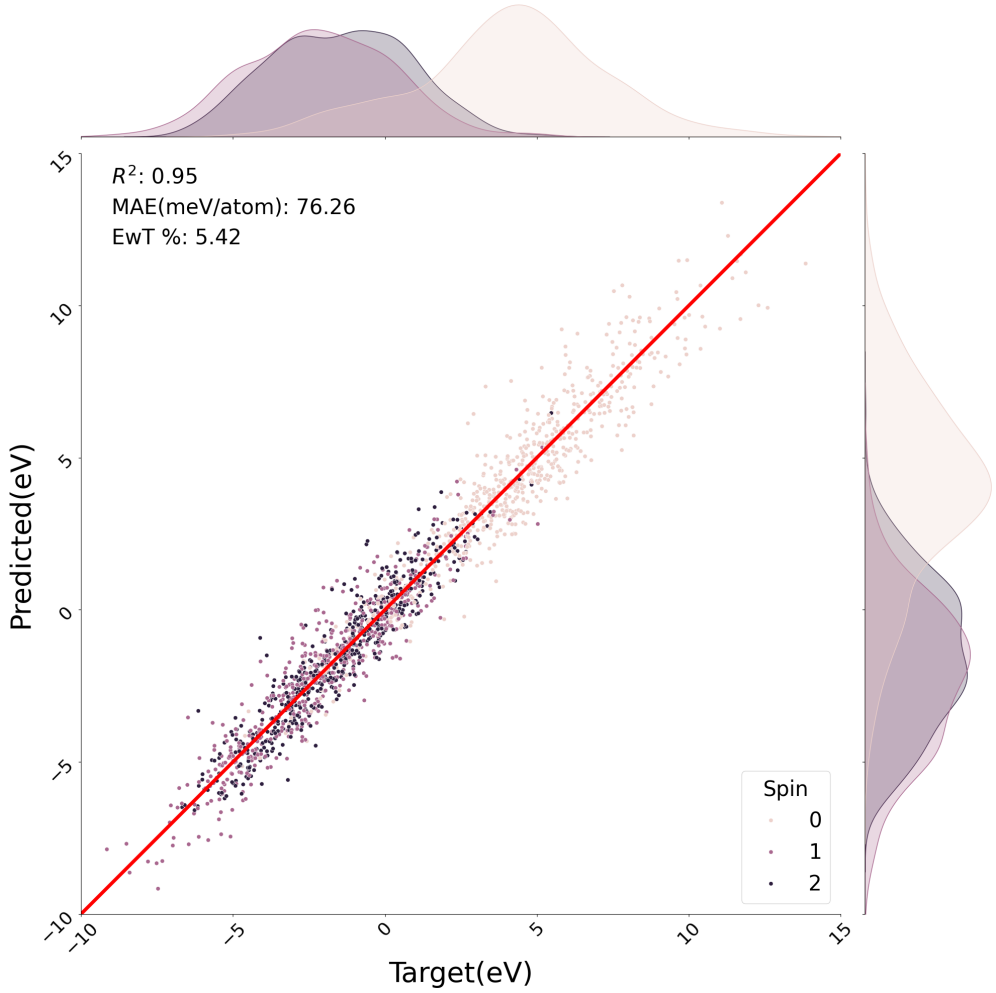

Figure S14: LIBE non-QTAIM test Partity.

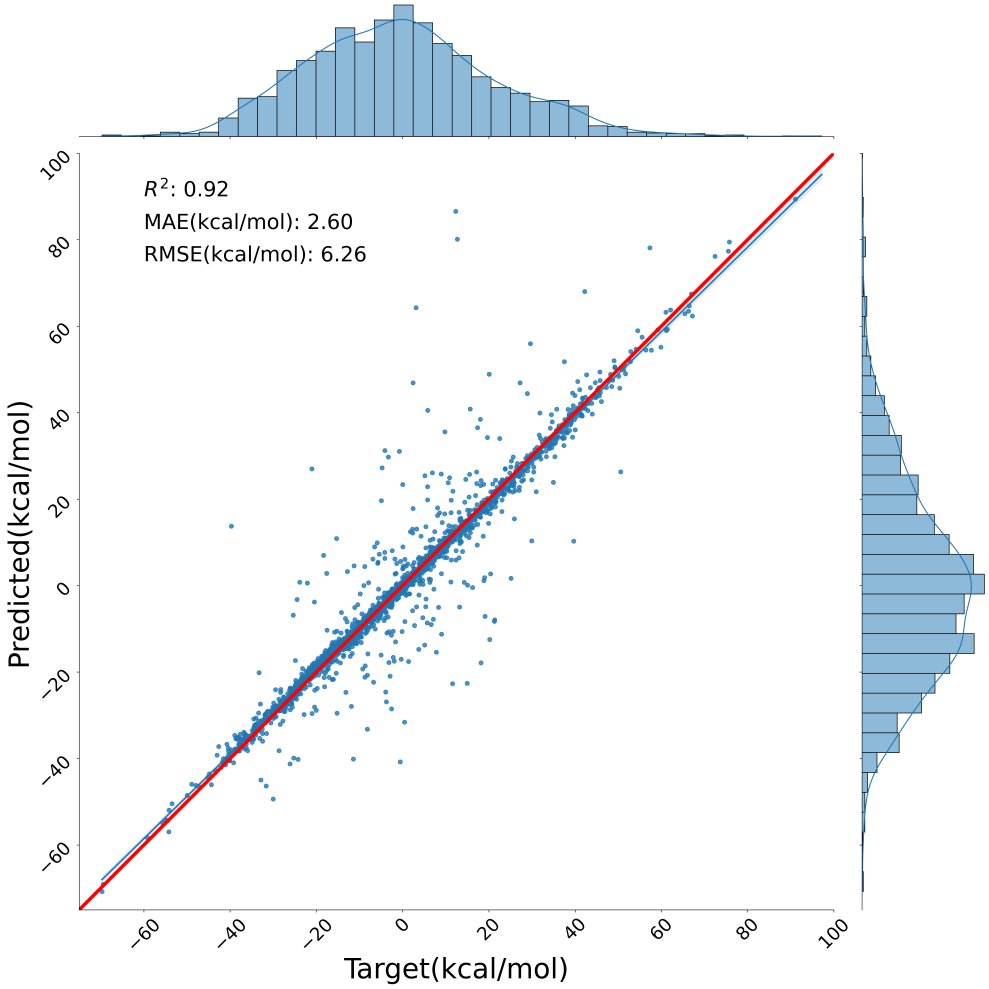

Figure S15: Green QTAIM test Partity.

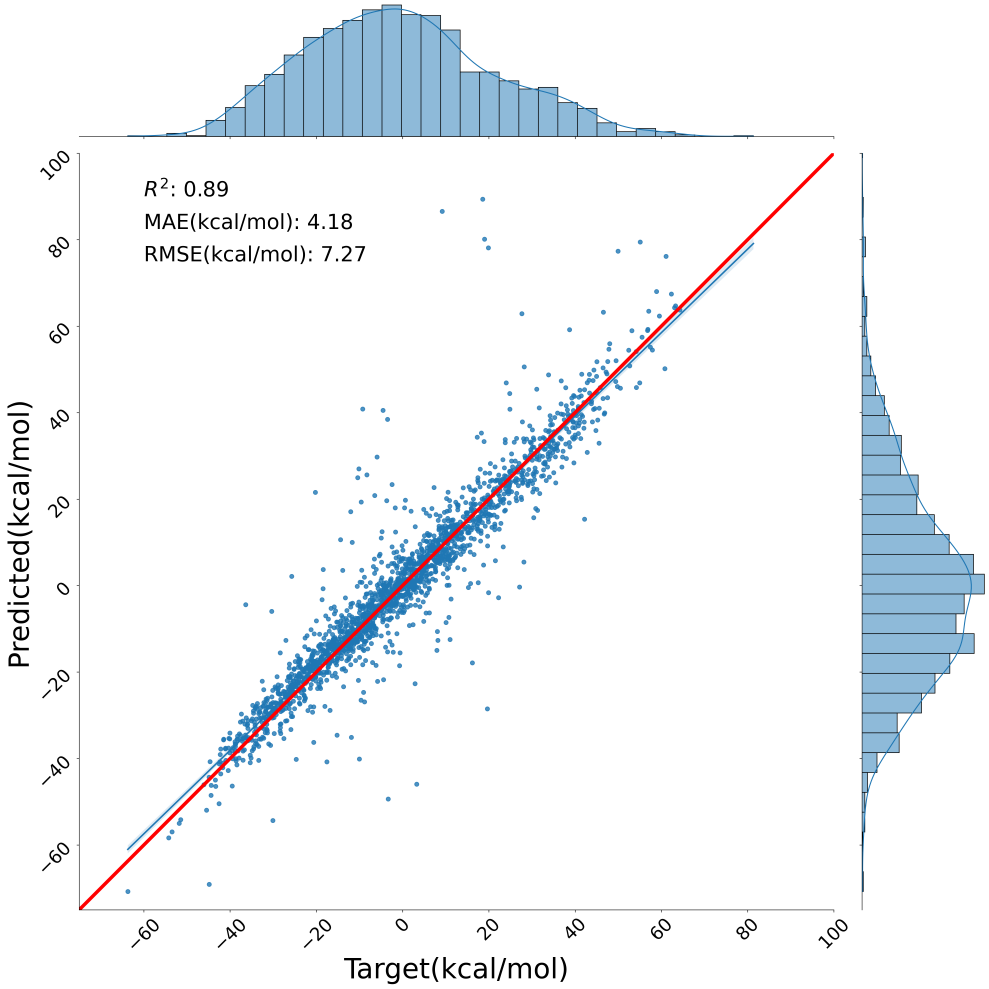

Figure S16: Green non-QTAIM test Partity.

OOD True vs. Predicted Plots

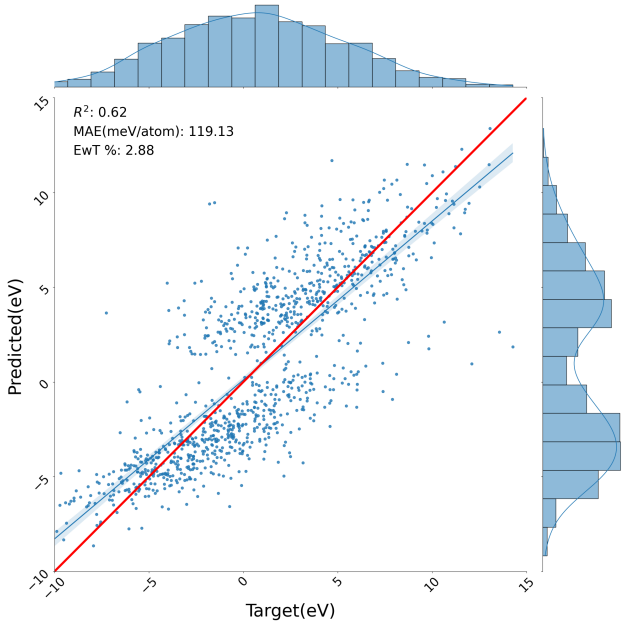

Figure S17: LIBE OOD QTAIM overall test Partity.

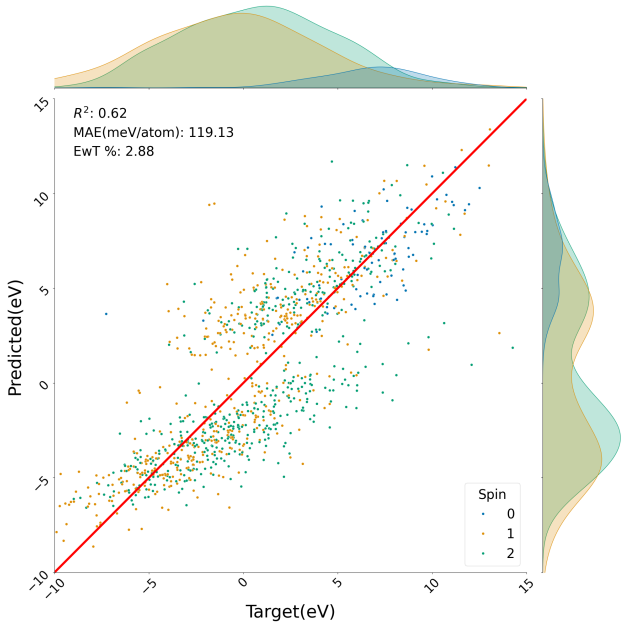

Figure S18: LIBE OOD QTAIM spin-stratified test Partity.

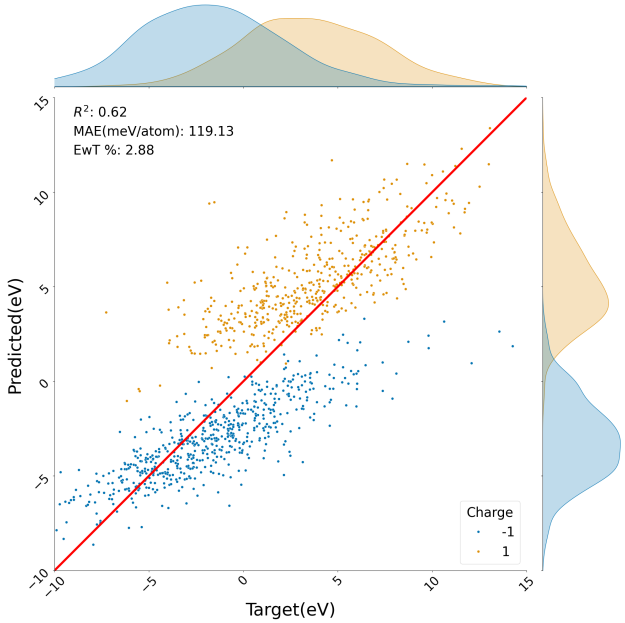

Figure S19: LIBE OOD QTAIM charge-stratified test Partity.

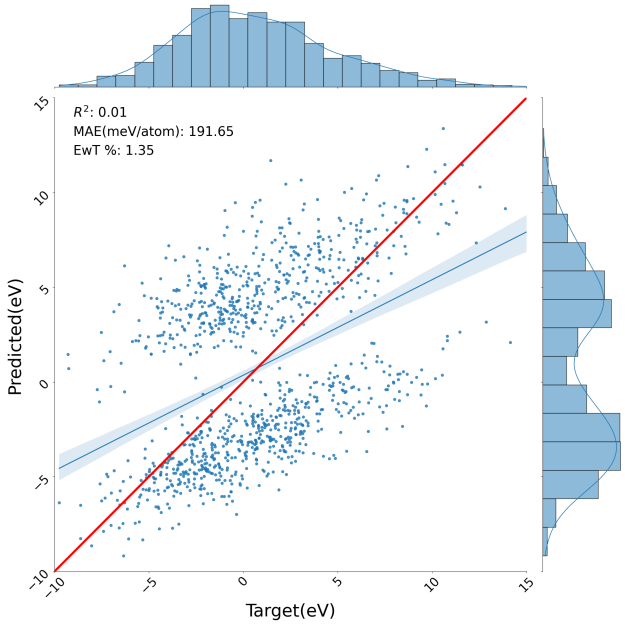

Figure S20: LIBE OOD non-QTAIM stratified test Partity.

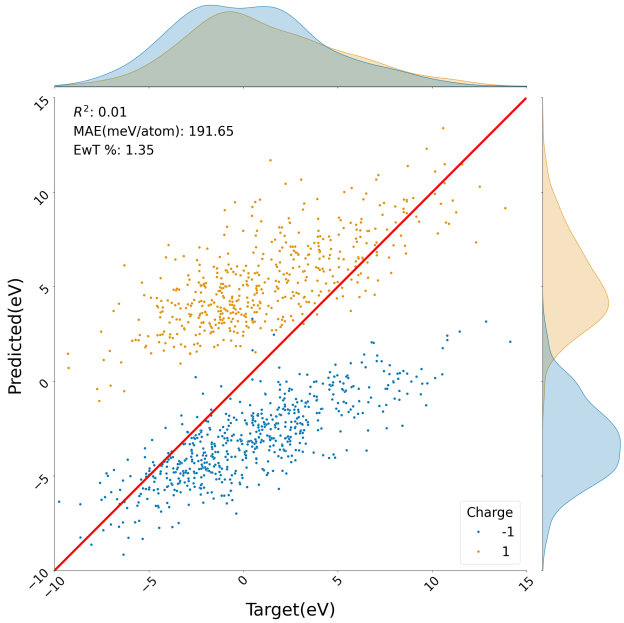

Figure S21: LIBE OOD non-QTAIM charge-stratified test Partity.

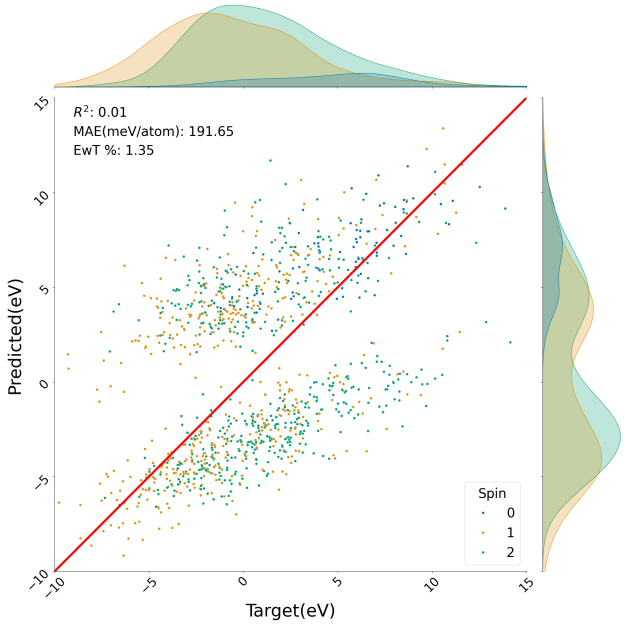

Figure S22: LIBE OOD non-QTAIM spin-stratified test Partity.

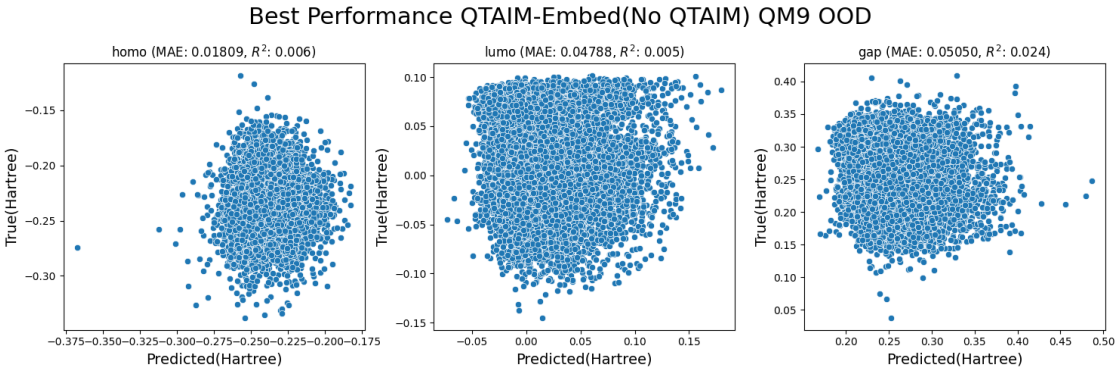

Figure S23: QM9 OOD non-QTAIM test Partity.

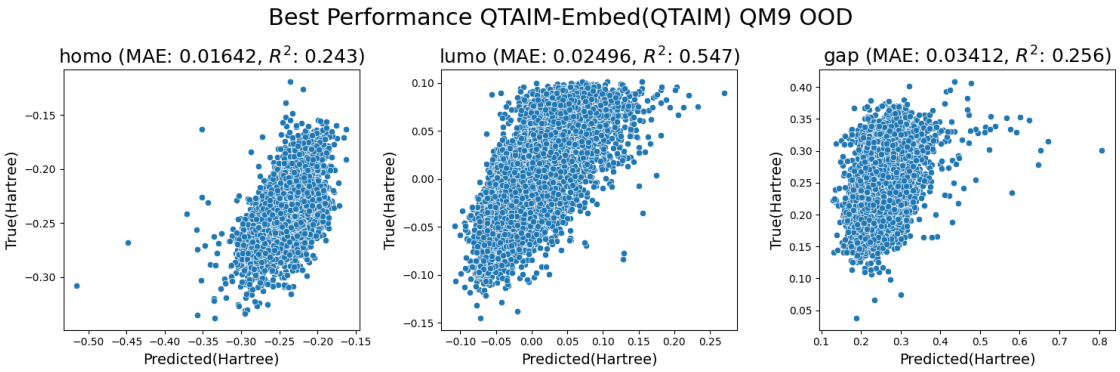

Figure S24: QM9 OOD QTAIM test Partity.

Tox21 Results

Table S25: Tox21 Test Performance

	Our Model (QTAIM)	**Our Model (No QTAIM)**
NR-AR	**0.9722**	0.9644
NR-AR-LBD	**0.9797**	0.9734
NR-AhR	**0.8899**	0.8824
NR-Aromatase	**0.9584**	0.9502
NR-ER	**0.8988**	0.8942
NR-ER-LBD	**0.9613**	0.9567
NR-PPAR-gamma	0.9779	**0.9786**
SR-ARE	**0.8567**	0.8413
SR-ATAD5	**0.9775**	0.9748
SR-HSE	**0.9506**	0.9467
SR-MMP	**0.8552**	0.8405
SR-p53	**0.9477**	0.9458
Average AUROC	**0.9355**	0.9291

Full Learning Curves

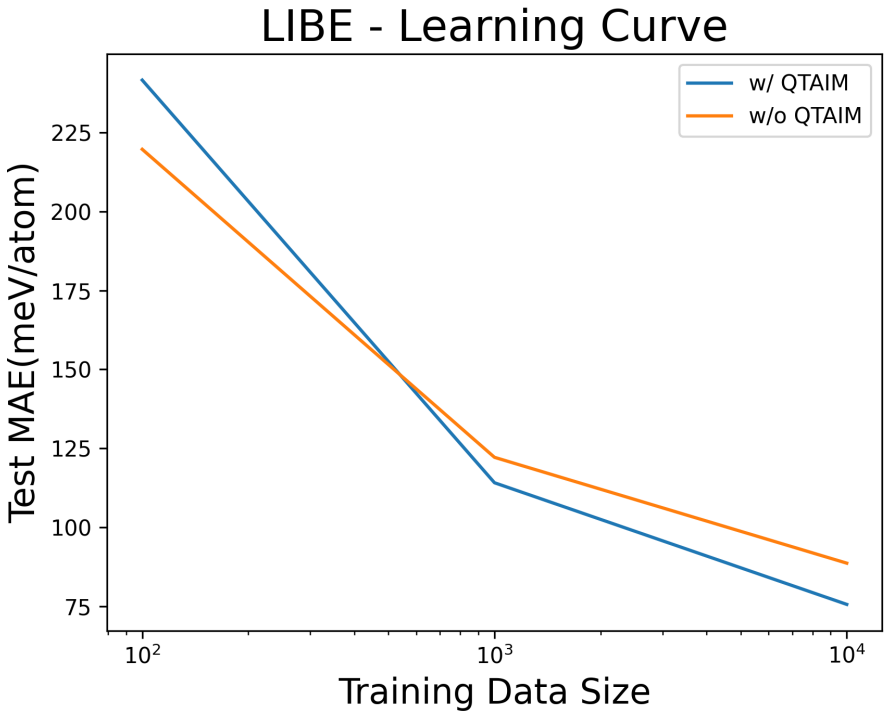

Figure S26: LIBE Learning Curve on MAE

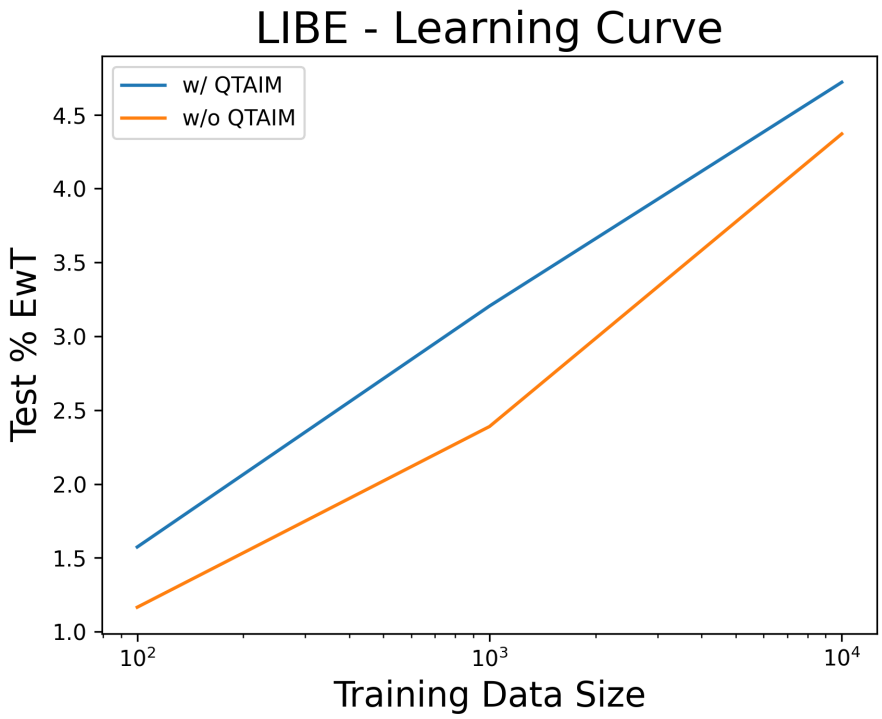

Figure S27: LIBE Learning Curve on %EwT

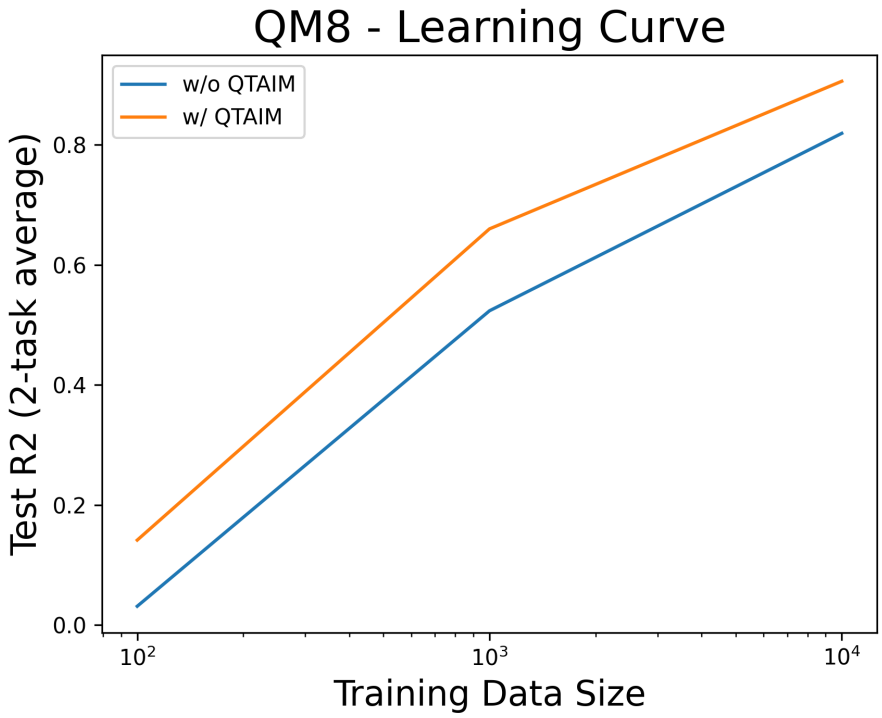

Figure S28: QM8 Learning Curve on R^2

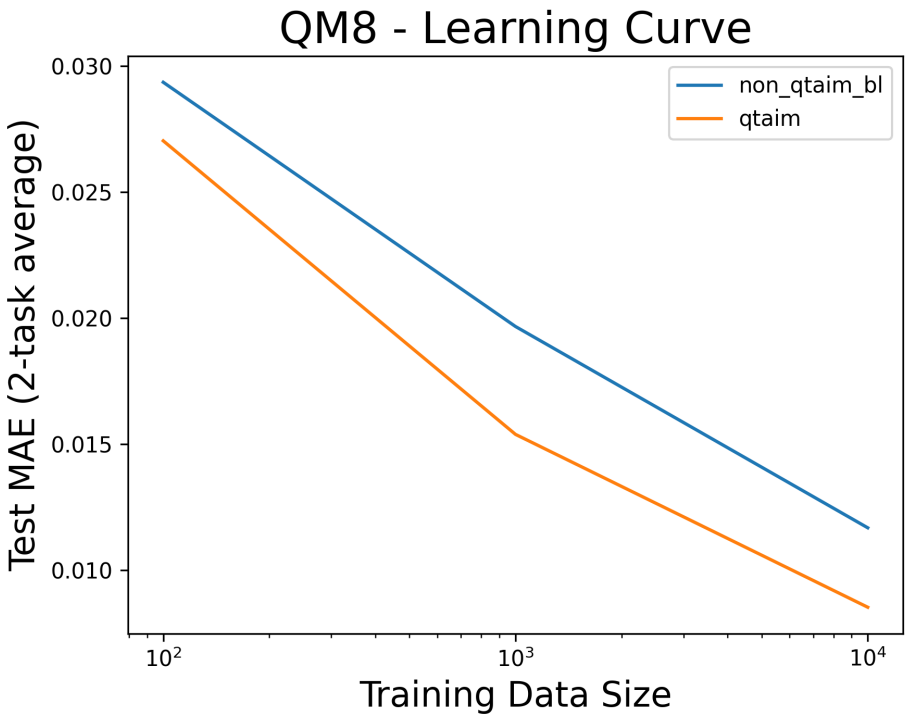

Figure S29: QM8 Learning Curve on MAE

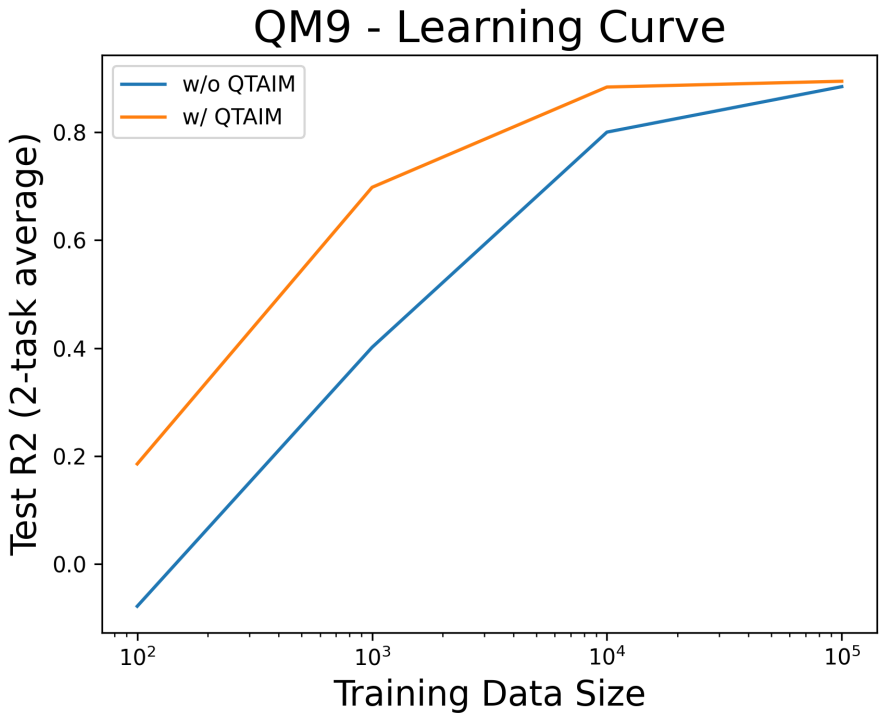

Figure S30: QM9 Learning Curve on R^2

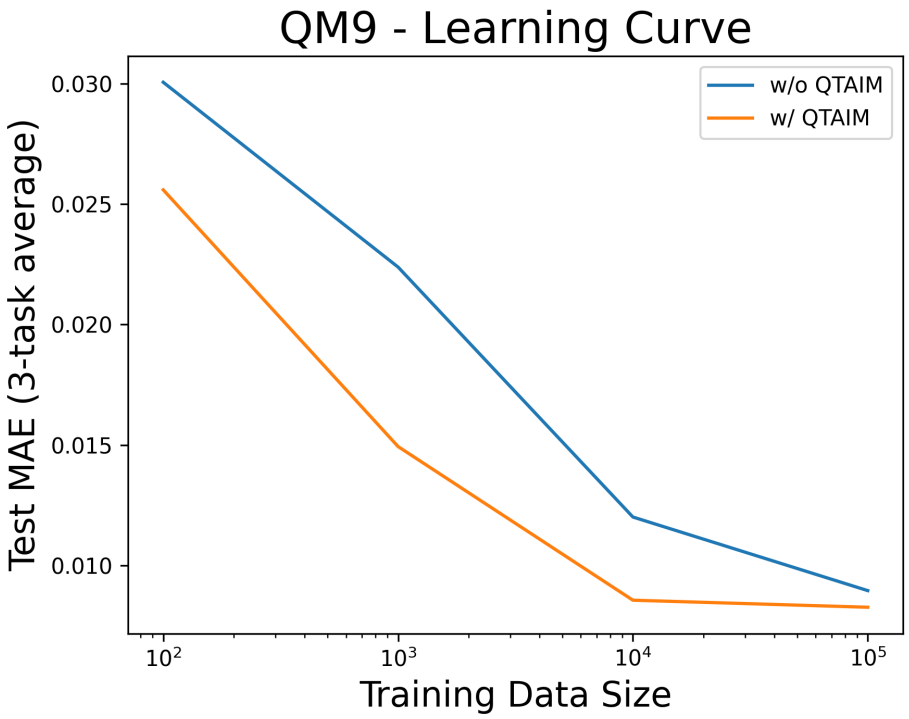

Figure S31: QM9 Learning Curve on MAE

Hyperparameter Selection

For ChemProp hyperparameter optimizations we used their inbuilt hyperopt functionality.[1] Here we used the following set of parameters as sweep values:

Schnet/PaiNN - Here we used their default values[2,3] a select set of values:

Table S32: Hyperparameters Swept for SchNet and PaiNN

Hyperparameter	**Values Swept**
N_atom_basis	10, 20, 50
Shared interaction	T F
LR	0.01, 0.001, 0.0001

QTAIM-embed - for our own in-house algorithms, we leveraged Wandb's parameter selection tool.[4] We used the same hyperparameter sweep configs for each model set. Finalized trained models for LIBE used the complete set of QTAIM descriptors above, other models removed α and β spin descriptors.

Table S33: QTAIM-embed (our) model hyperparameter sweeps

Hyperparameter	**Values Swept**
weight_decay	0.0, 0.00001
Embedding_size	16, 20, 24
Gated_dropout	0.0, 0.1, 0.2
Gated_hidden_size	64, 128
Gated_batch_norm	T, f
Gated_graph_norm	T, f
Num_lstm_iters	9, 11, 13, 15
Num_lstm_layers	1, 2
Fc_dropout	0.1, 0.2
Fc_hidden_size_1	256, 128
Fc_hidden_shape	flat, cone
Precision	bf16, 32
Gradient_clip_val	10, 100
Accumulated_grad_batches	1, 3, 5

Dataset QTAIM distributions

QM8

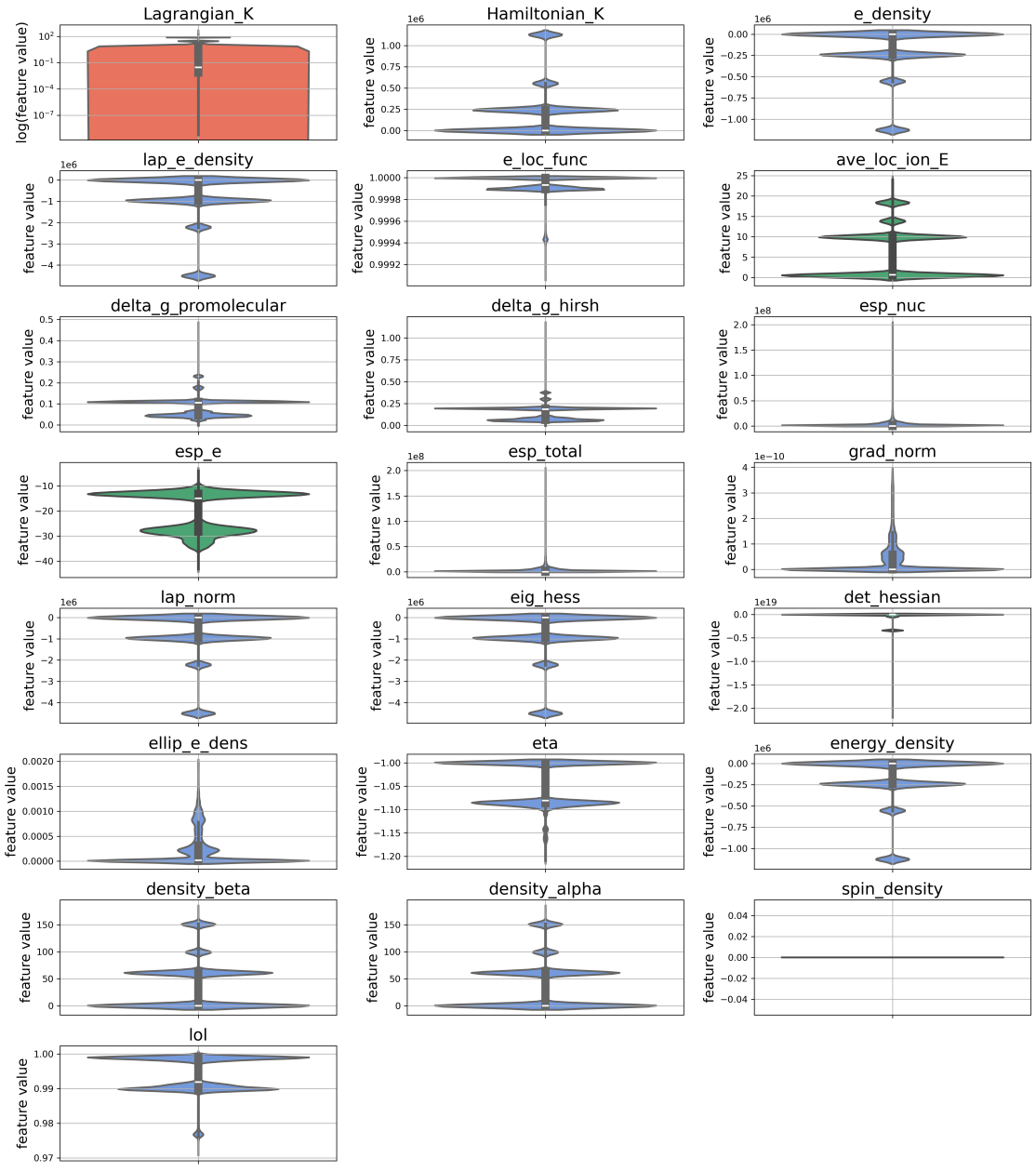

Figure S34: QM8 QTAIM value distribution - NCPs

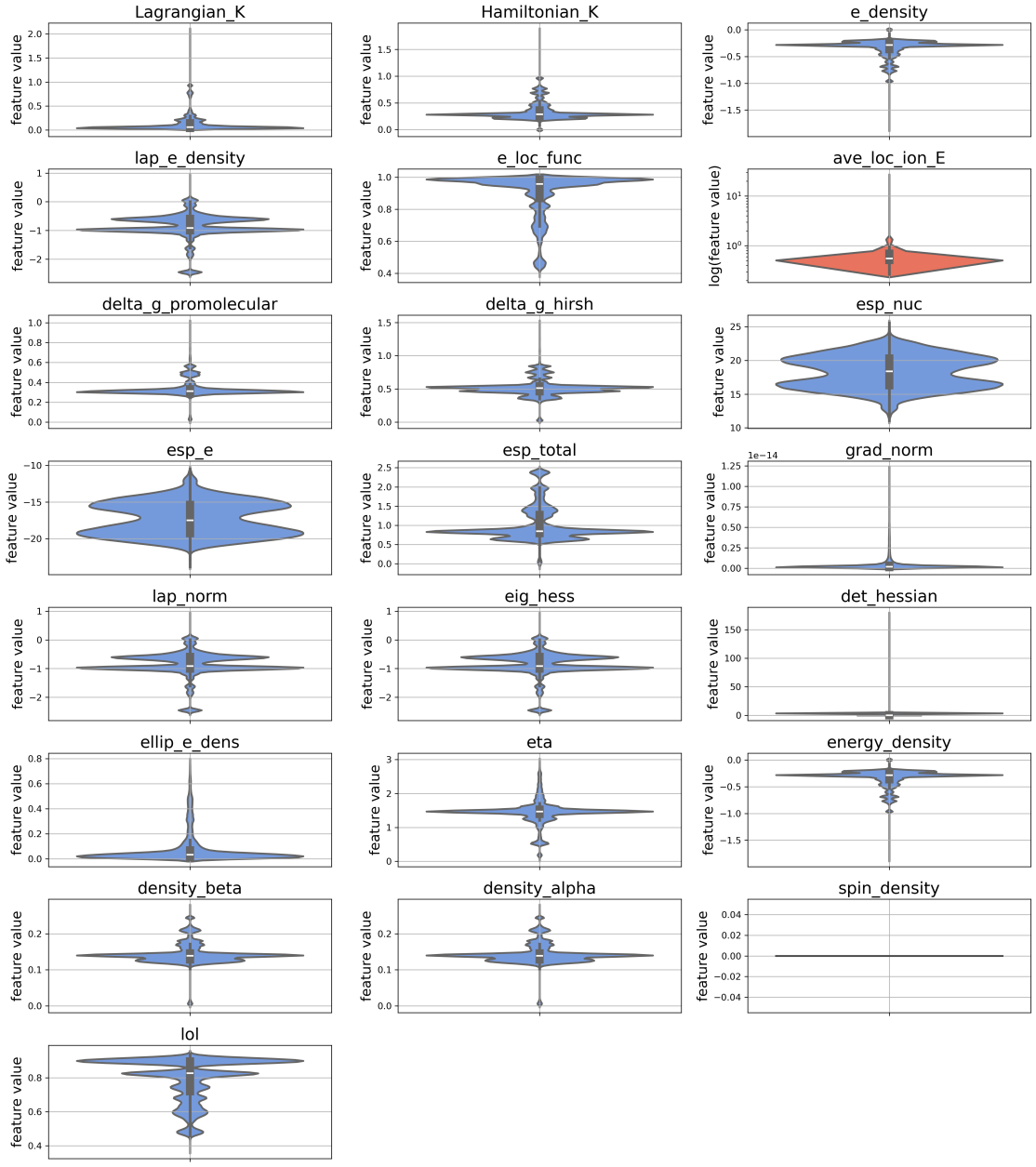

Figure S35: QM8 QTAIM value distribution - BCPs

QM9

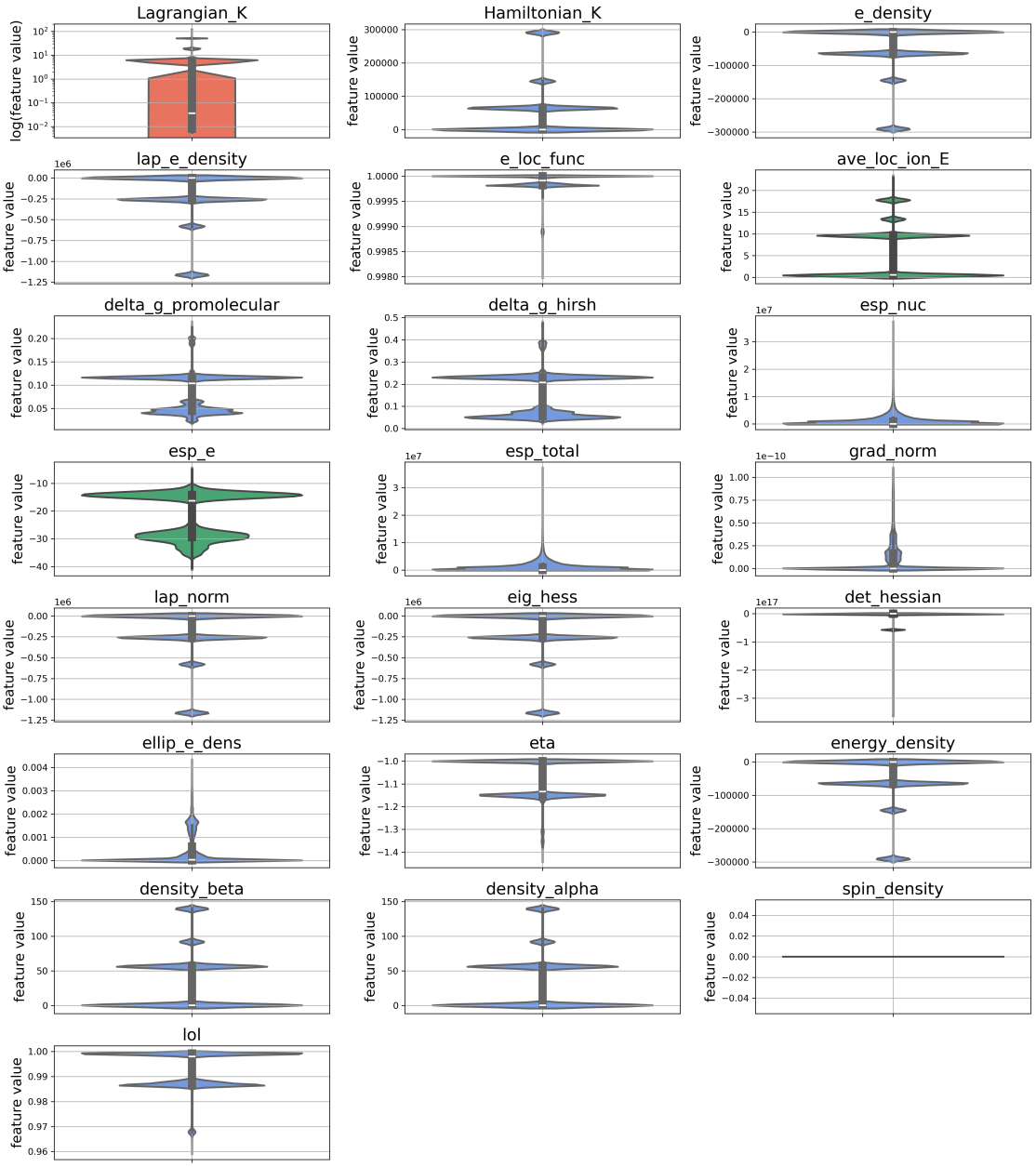

Figure S36: QM9 QTAIM value distribution - NCPs

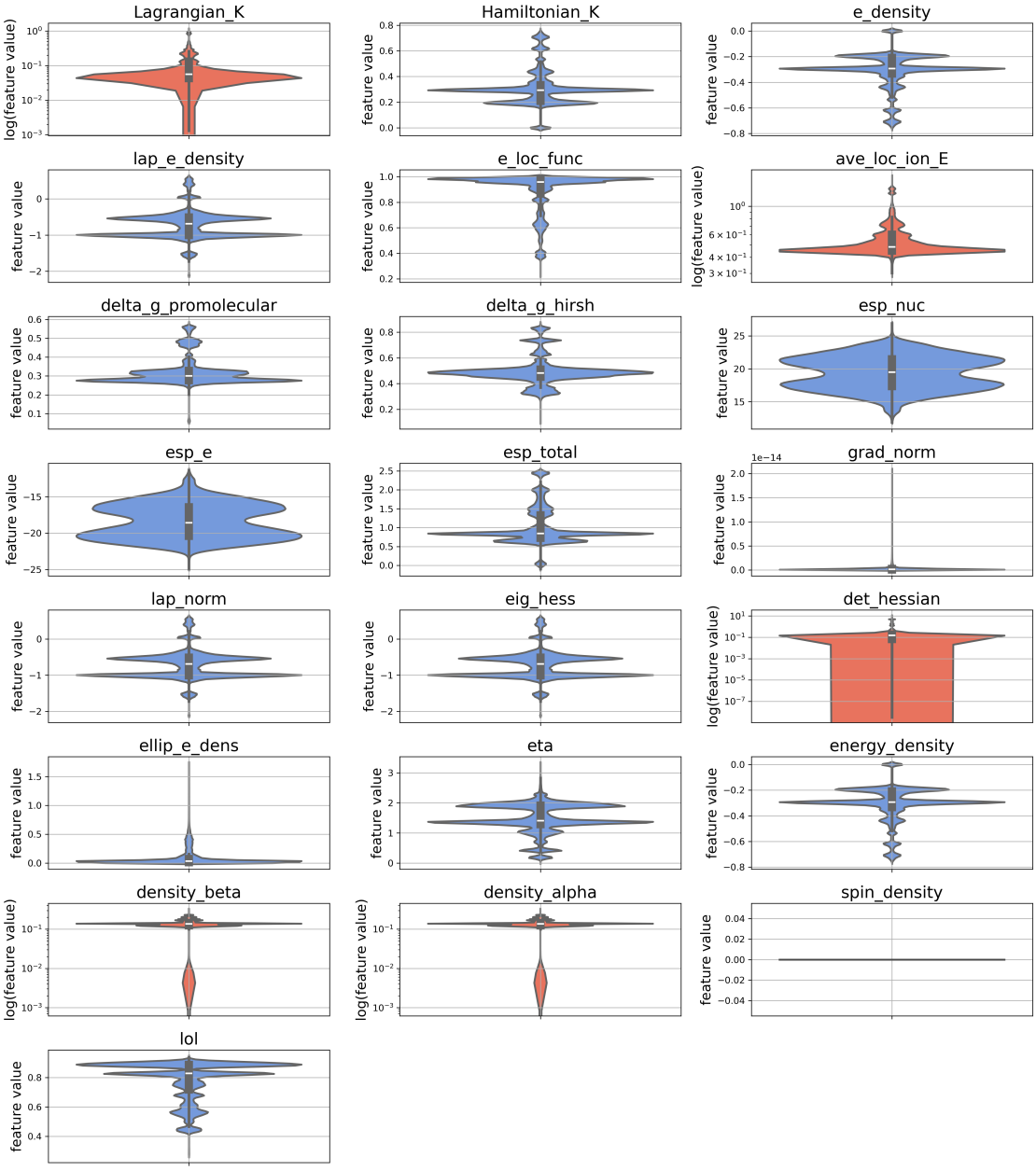

Figure S37: QM9 QTAIM value distribution - BCPs

LIBE

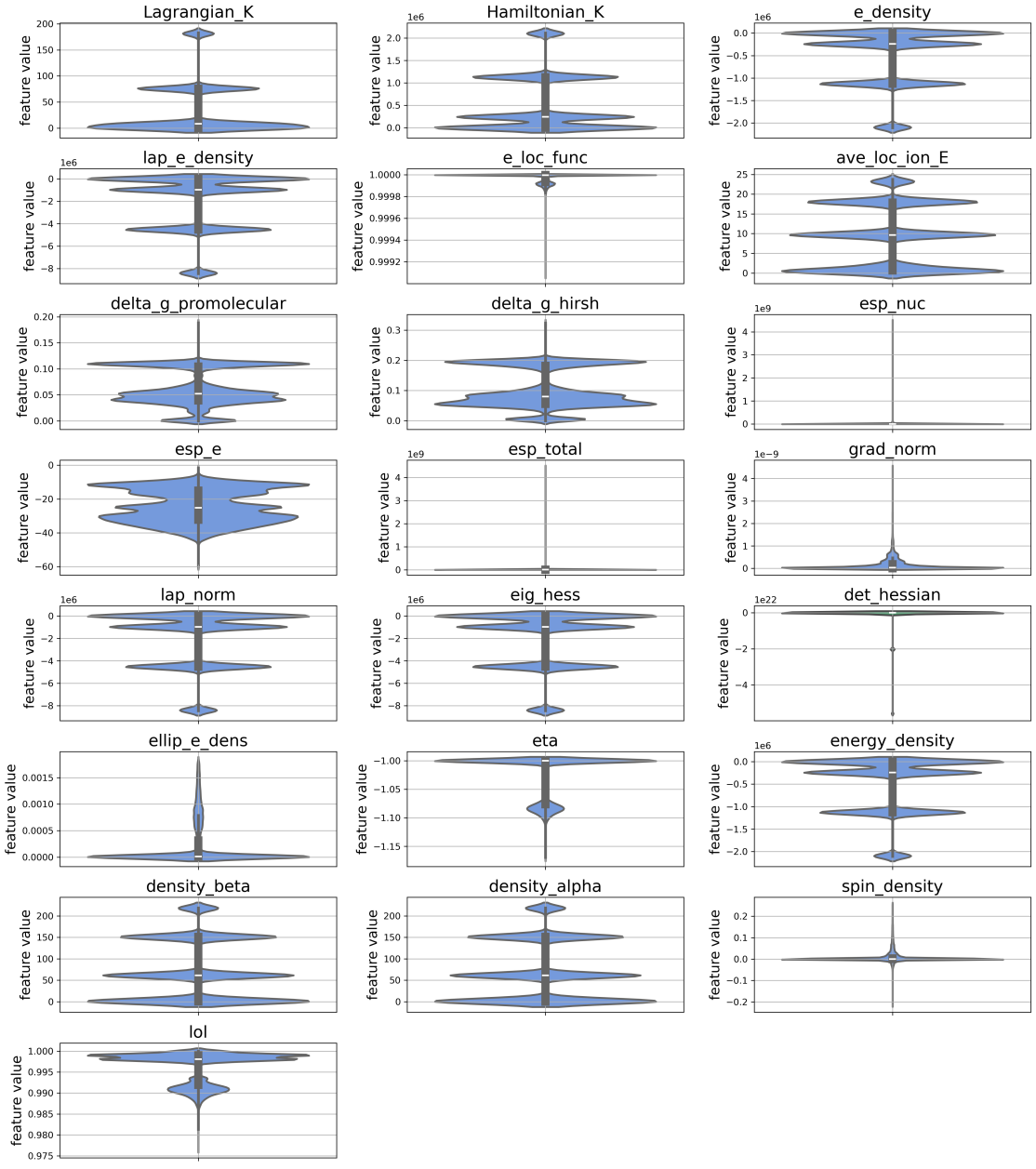

Figure S38: LIBE QTAIM value distribution - BCPs

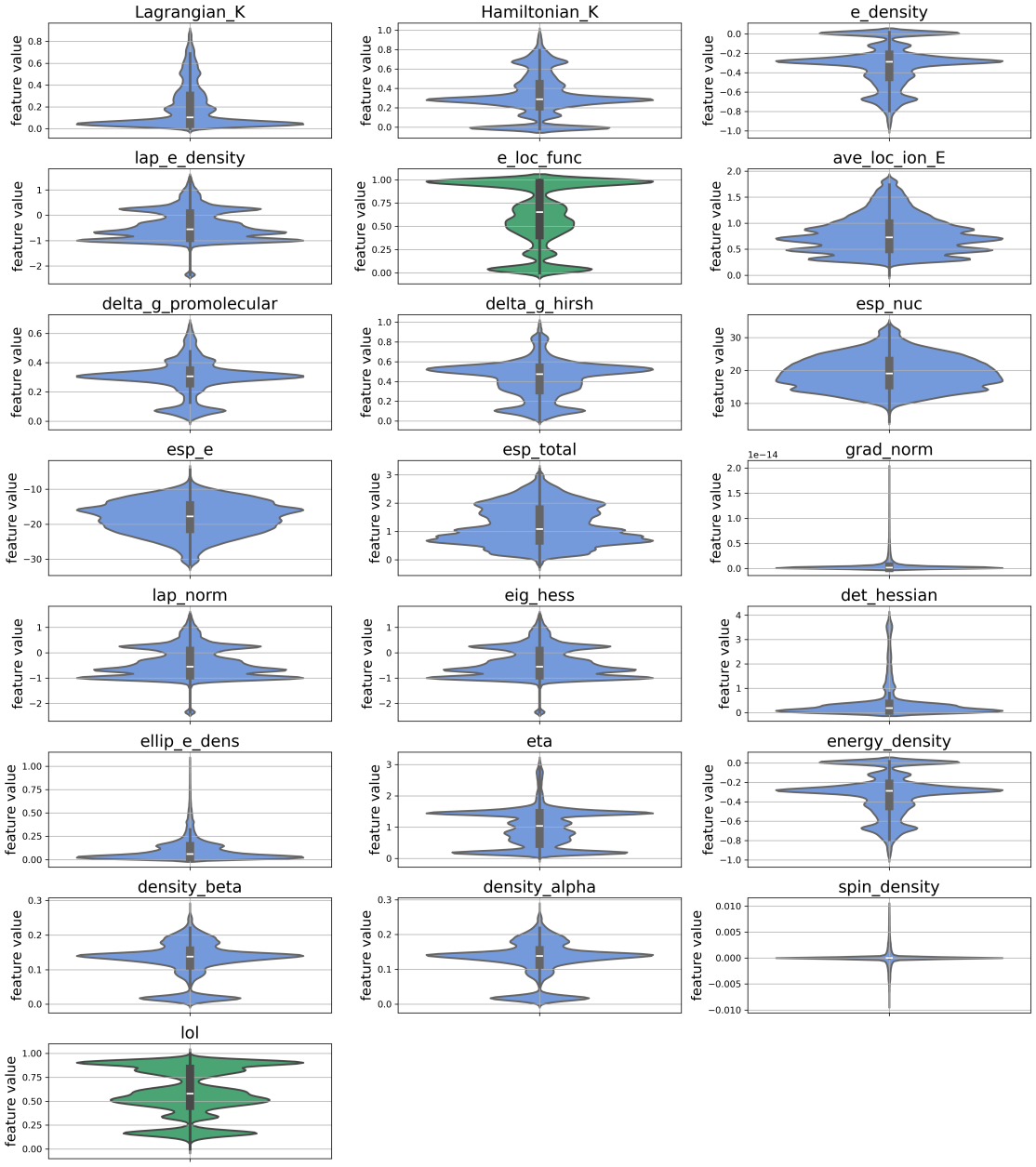

Figure S39: LIBE QTAIM value distribution - BCPs

Tox21

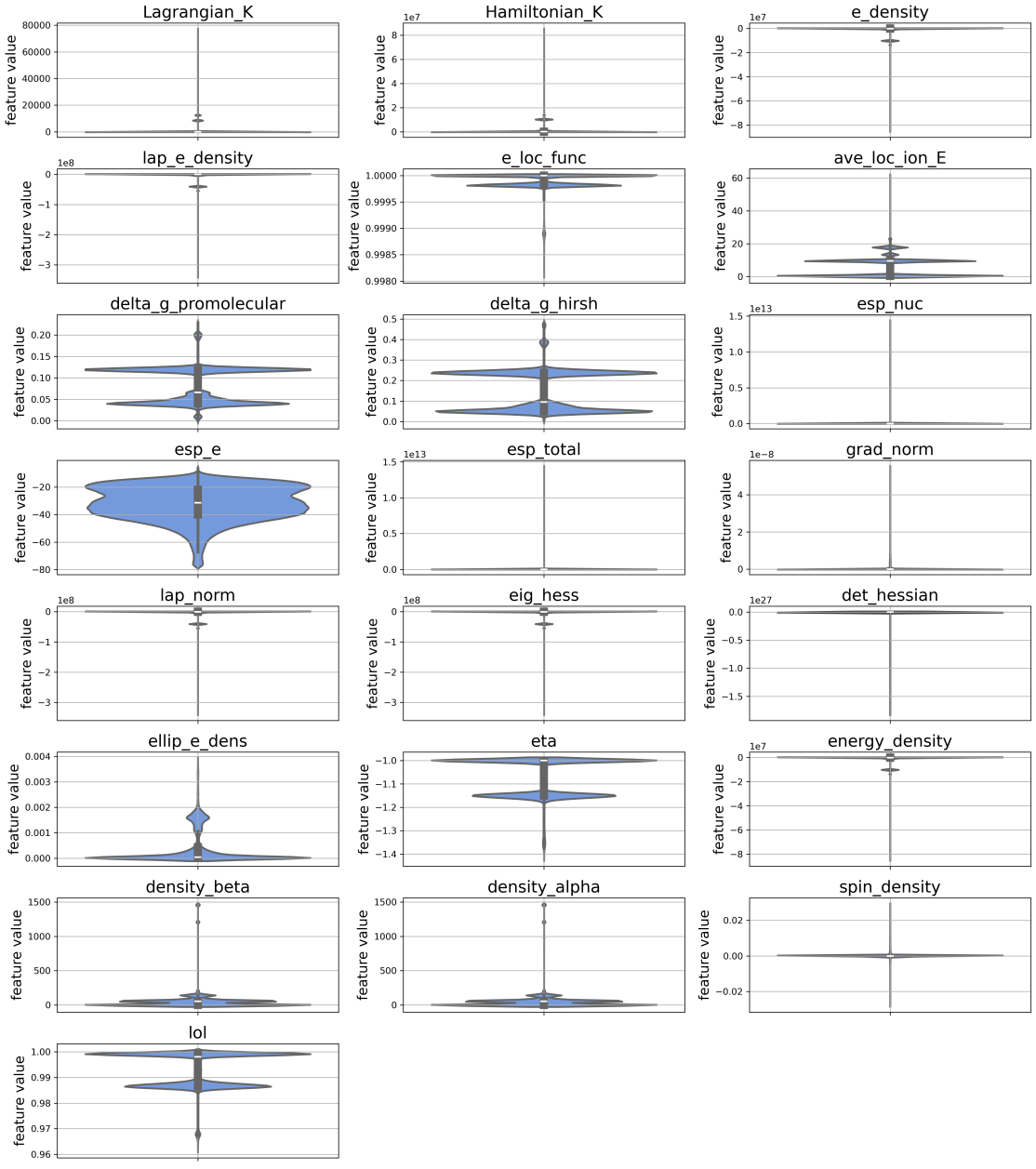

Figure S40: Tox21 QTAIM value distribution - BCPs

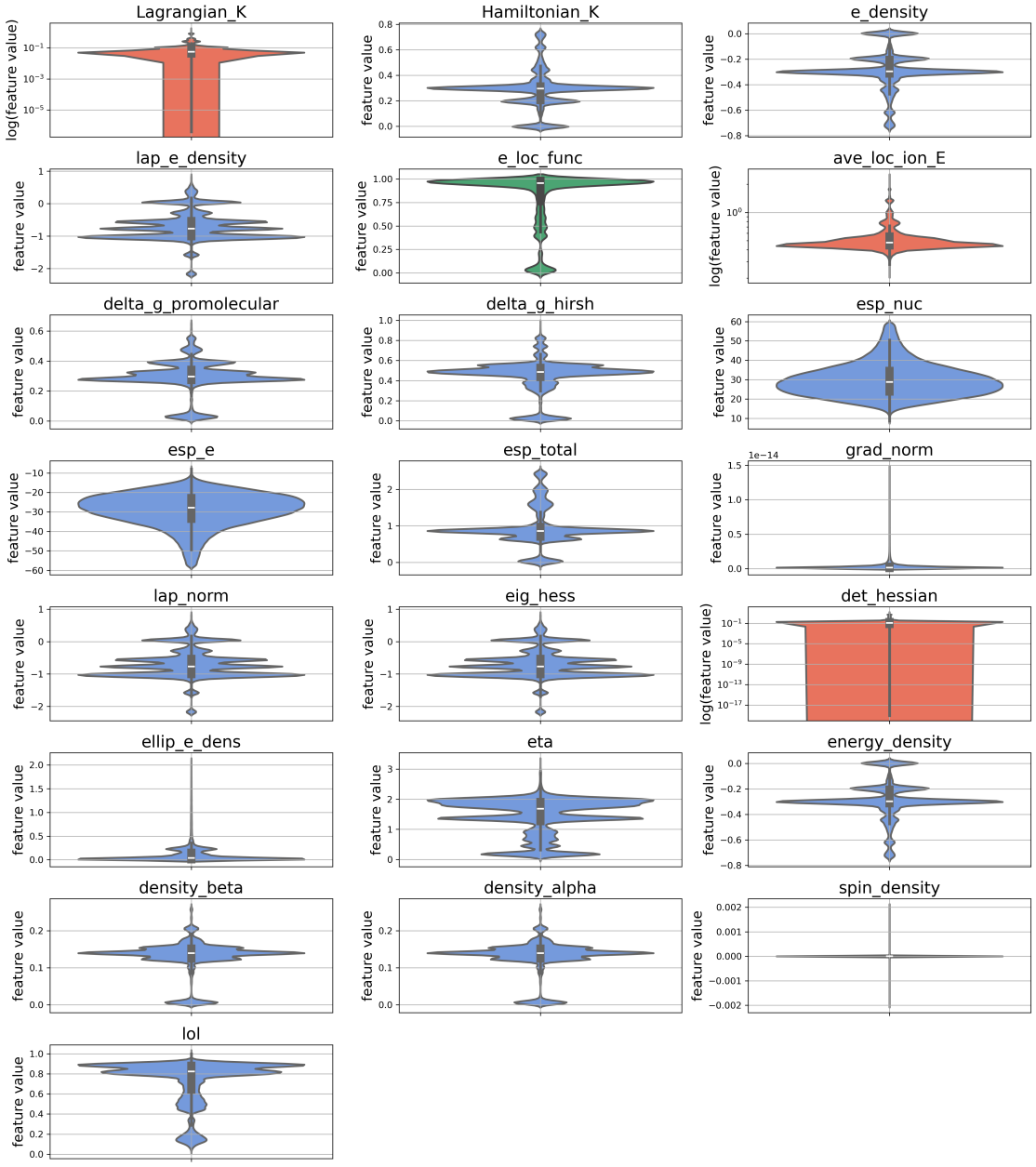

Figure S41: Tox21 QTAIM value distribution - BCPs

Correlation of QTAIM Values to Targets

QM8

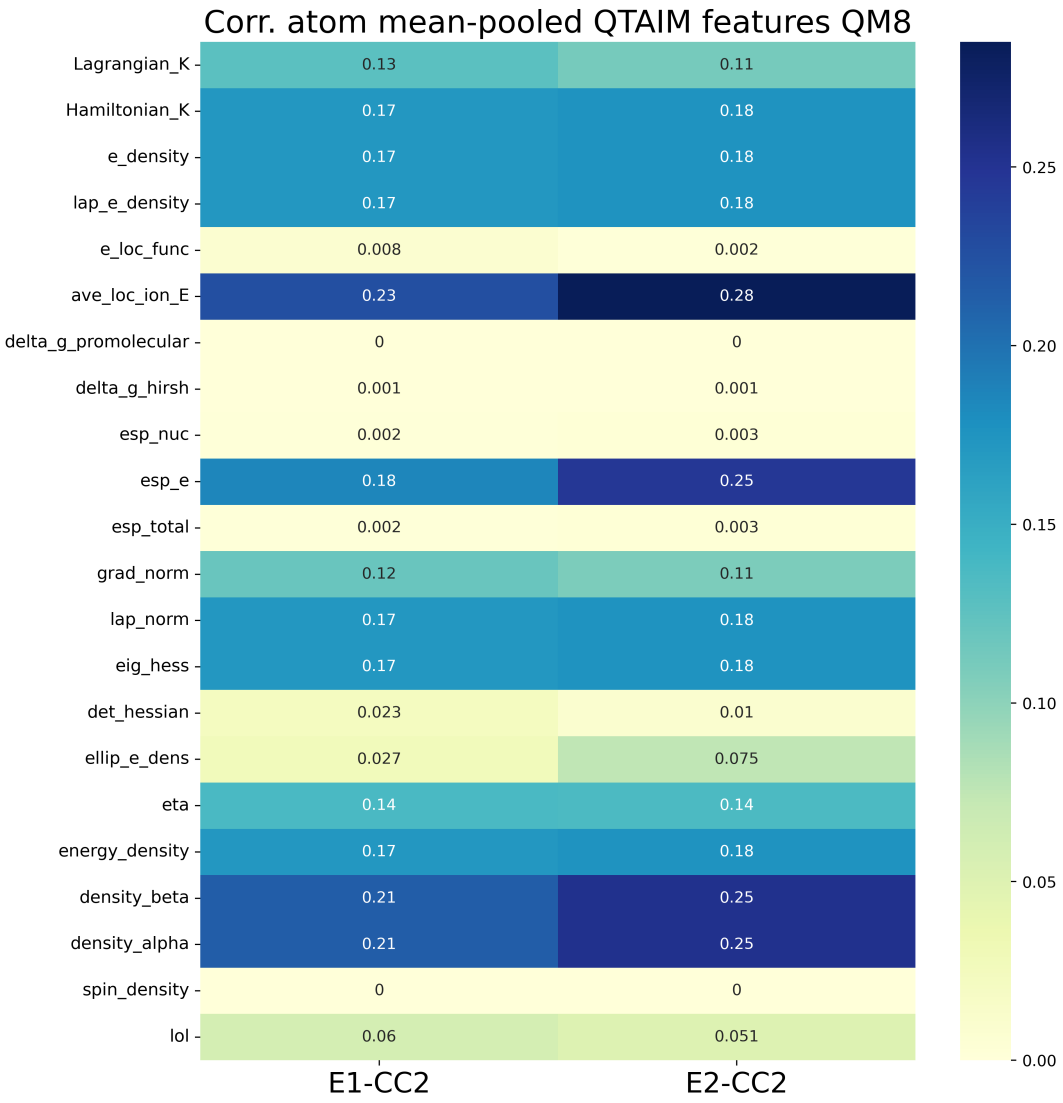

Figure S42: Correlation of NCP values with QM8 target values

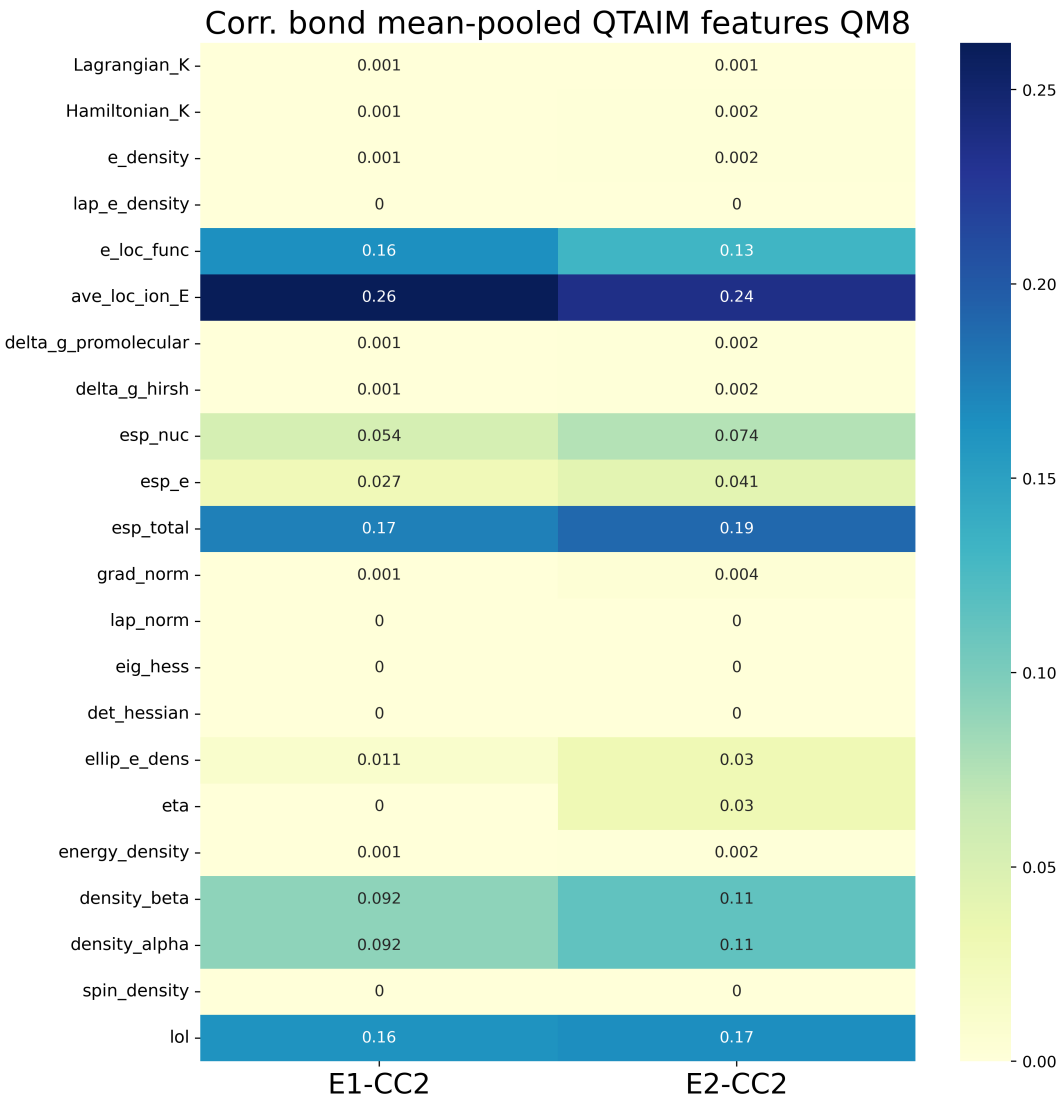

Figure S43: Correlation of BCP values with QM8 target values

QM9

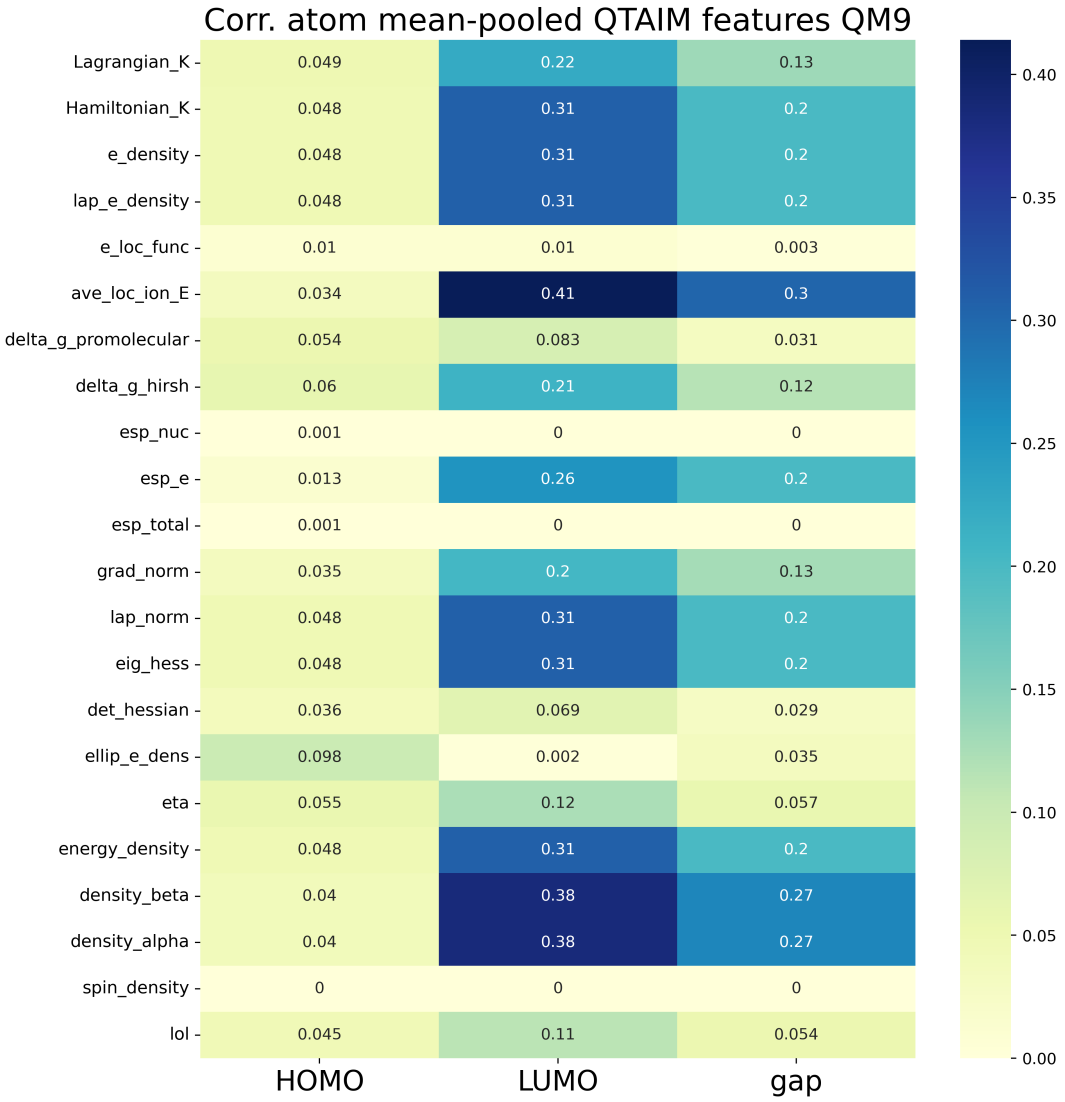

Figure S44: Correlation of NCP values with QM9 target values

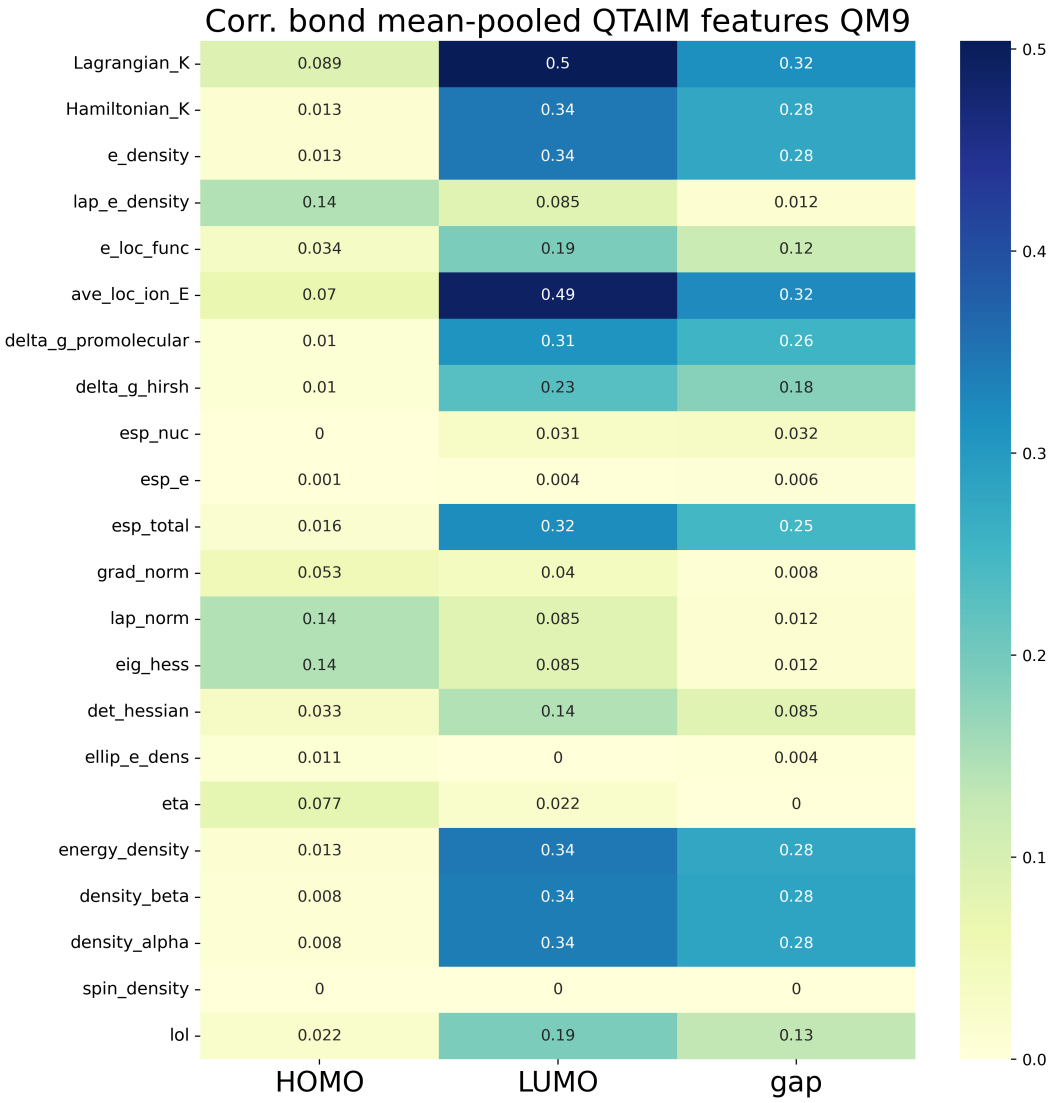

Figure S45: Correlation of BCP values with QM9 target values

LIBE

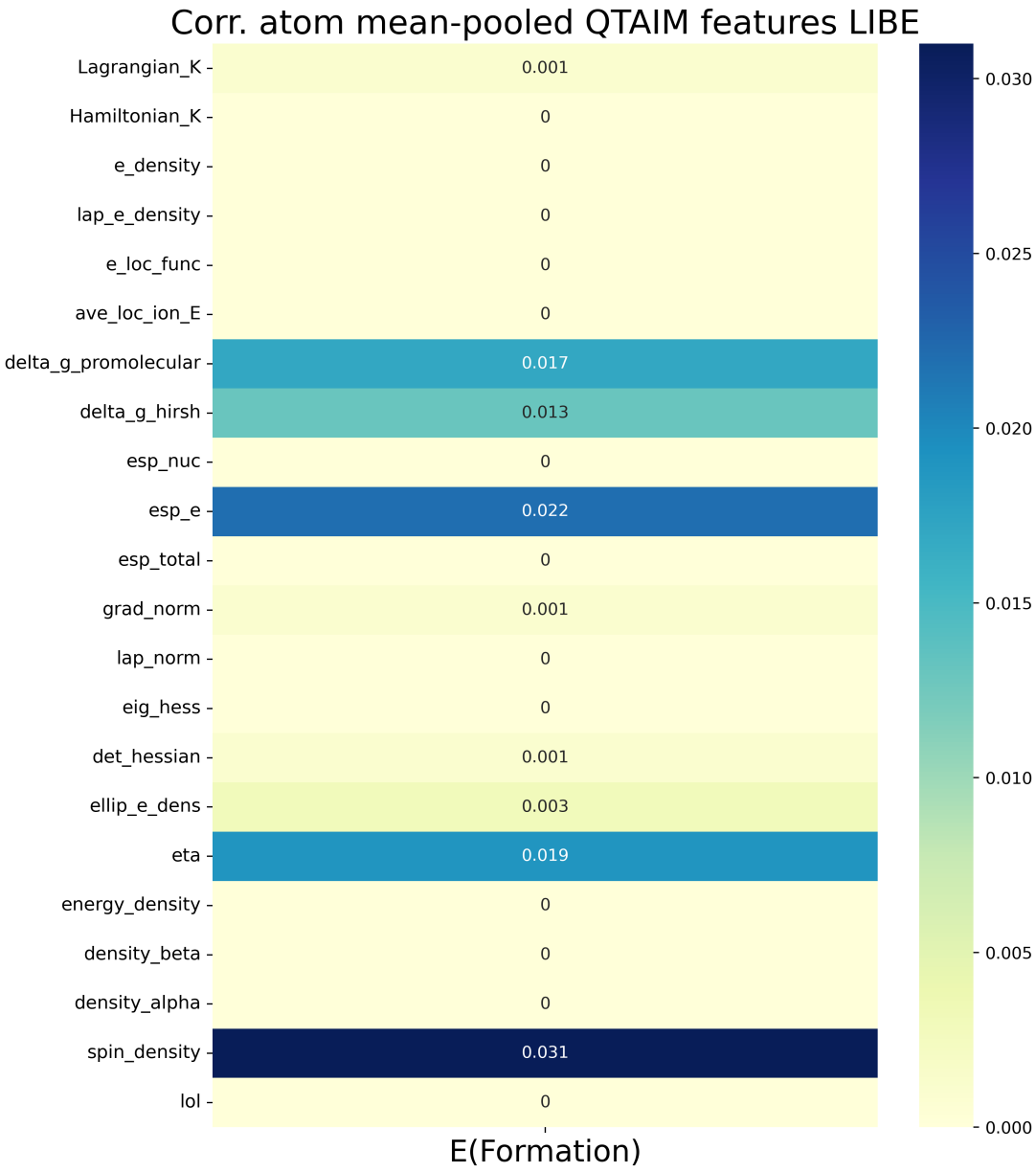

Figure S46: Correlation of NCP values with LIBE target values

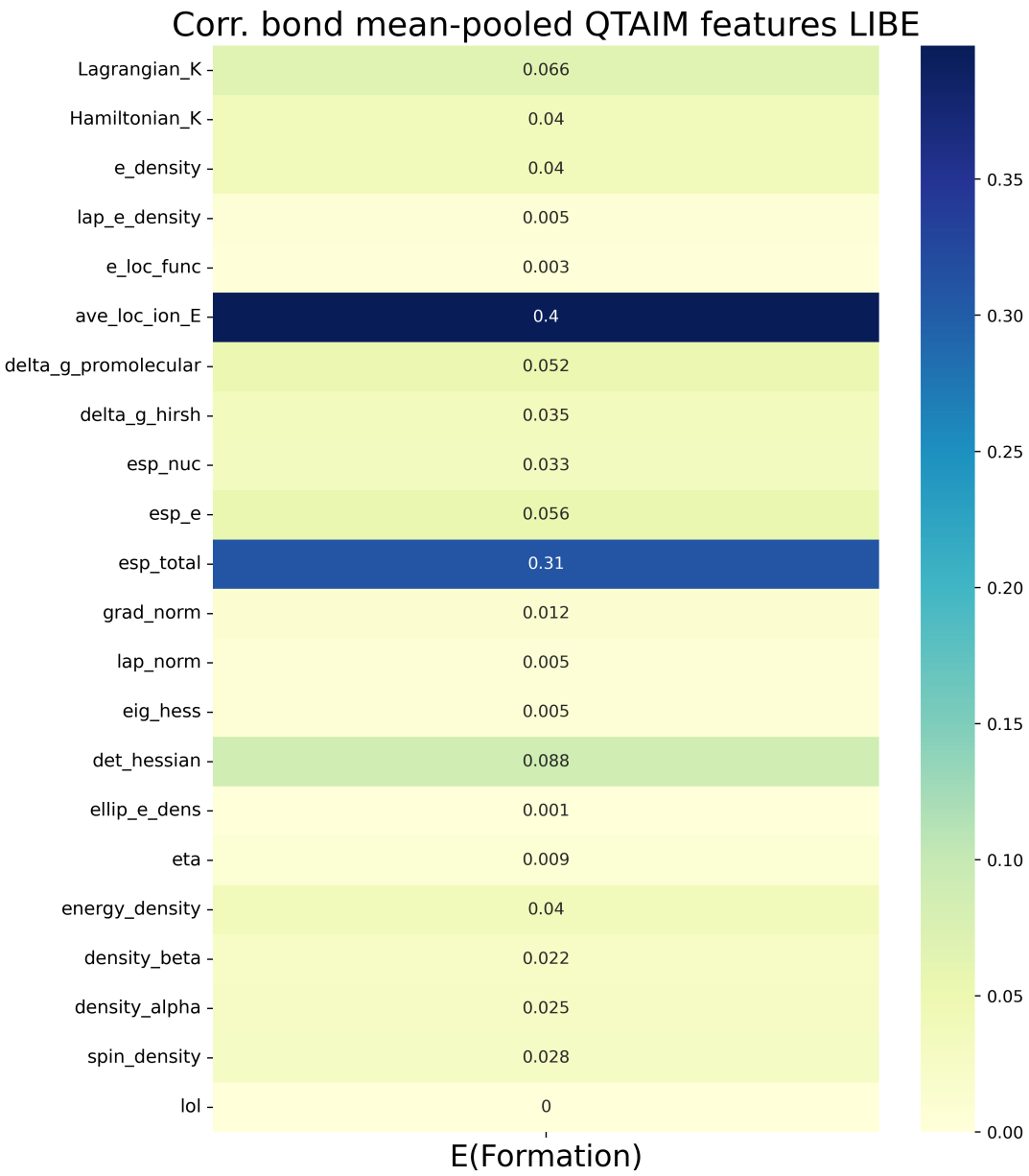

Figure S47: Correlation of BCP values with LIBE target values

Scatterplots of competing models

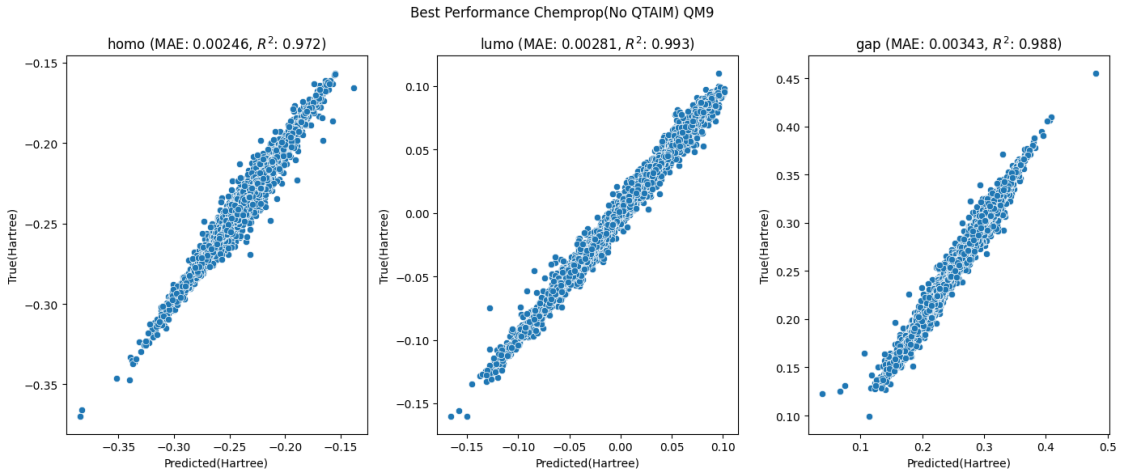

Figure S48: Parity Plot QM9 chemprop no QTAIM

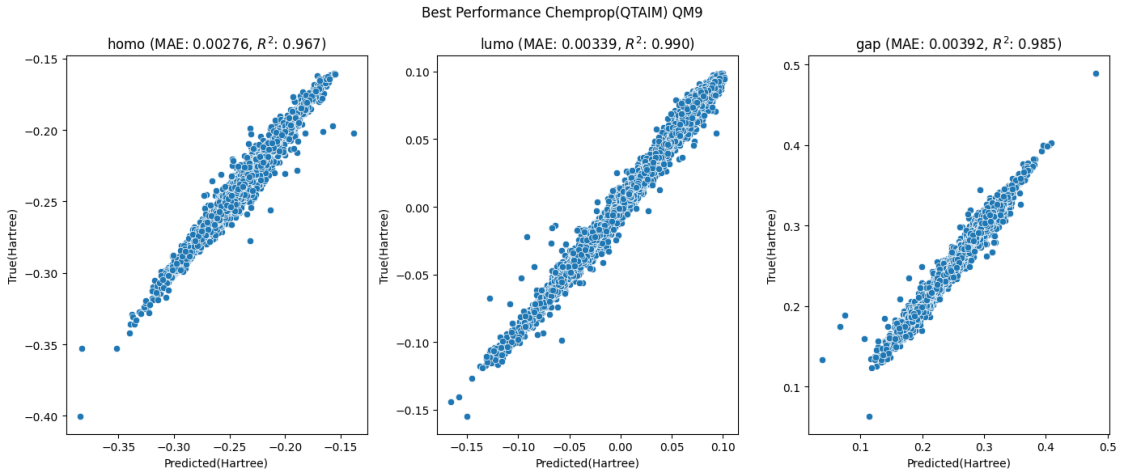

Figure S49: Parity Plot QM9 chemprop QTAIM

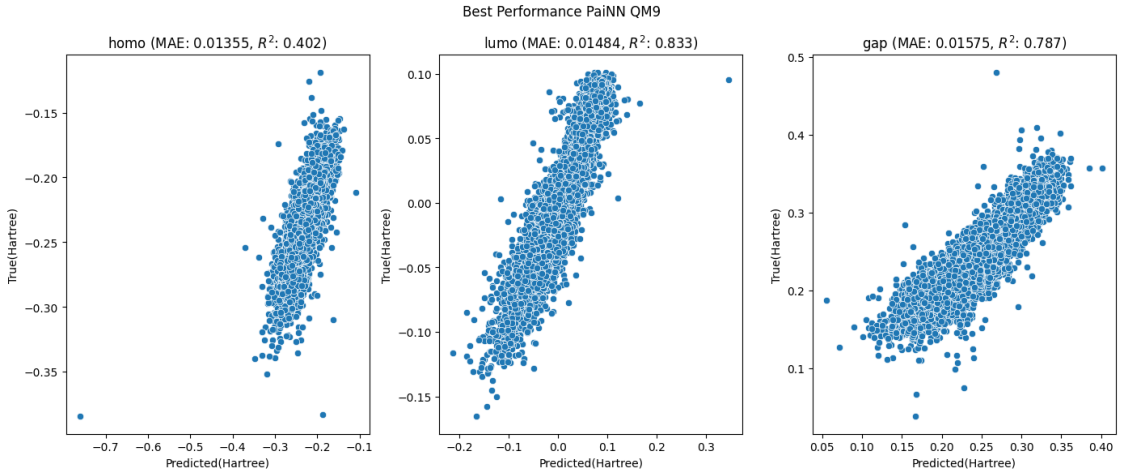

Figure S50: Parity Plot QM9 painn

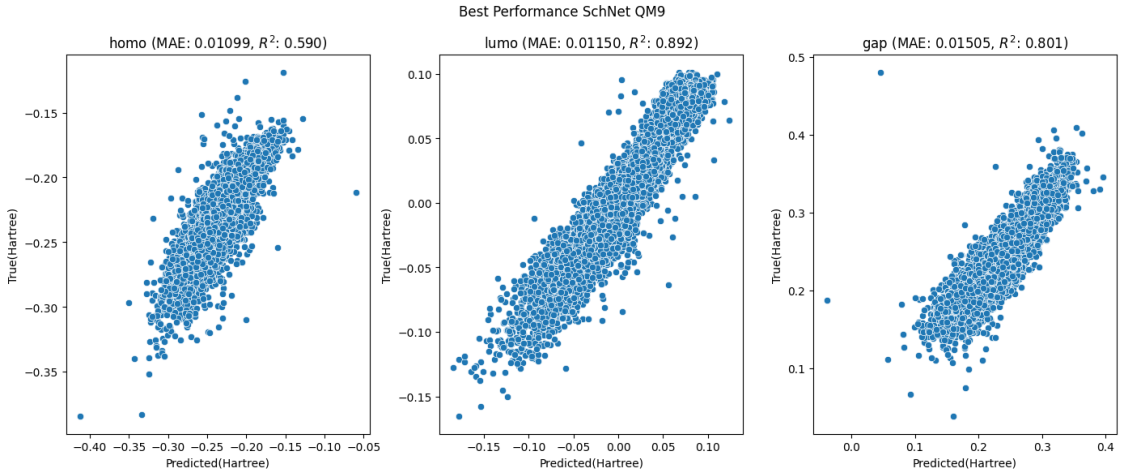

Figure S51: Parity Plot QM9 Schnet

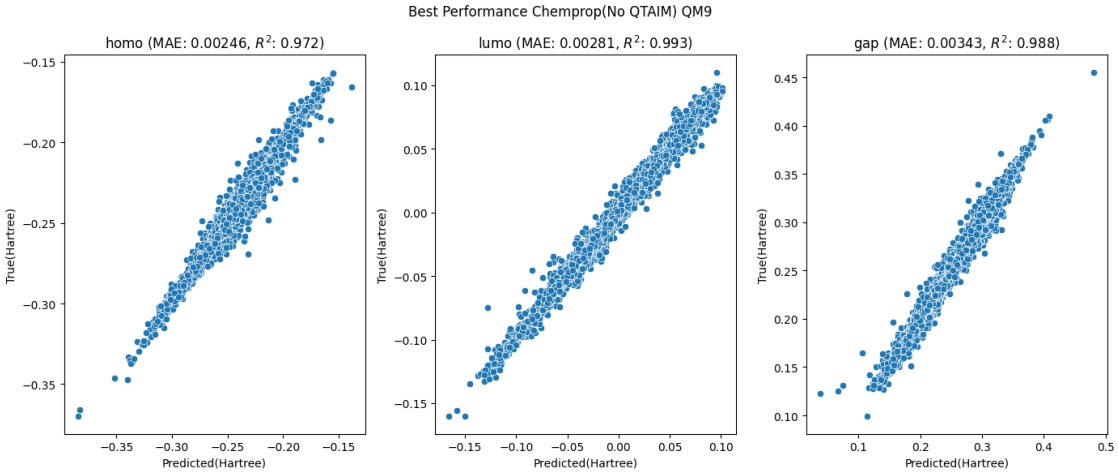

Figure S52: Parity Plot QM8 chemprop no QTAIM

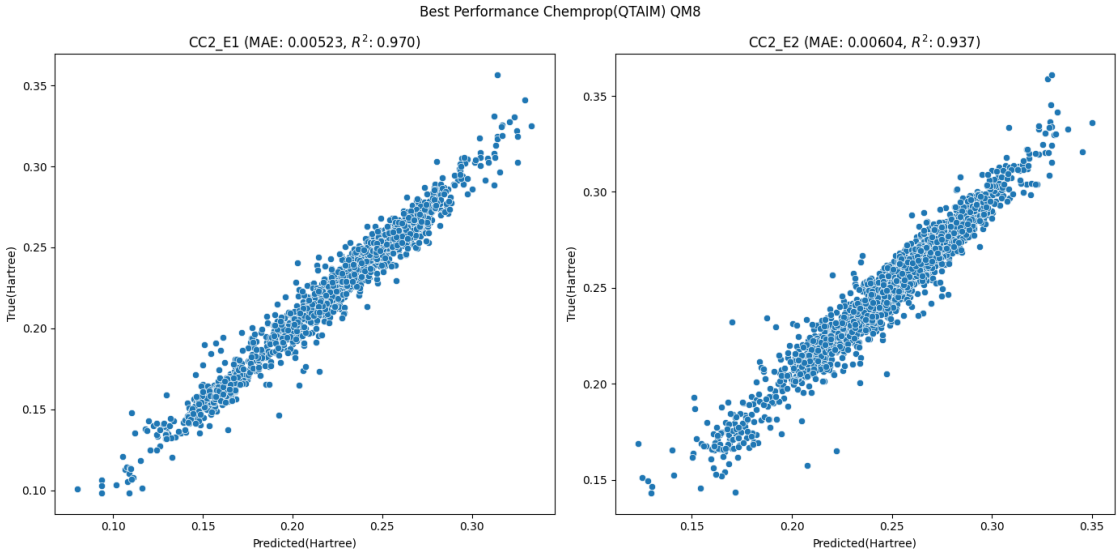

Figure S53: Parity Plot QM8 chemprop QTAIM

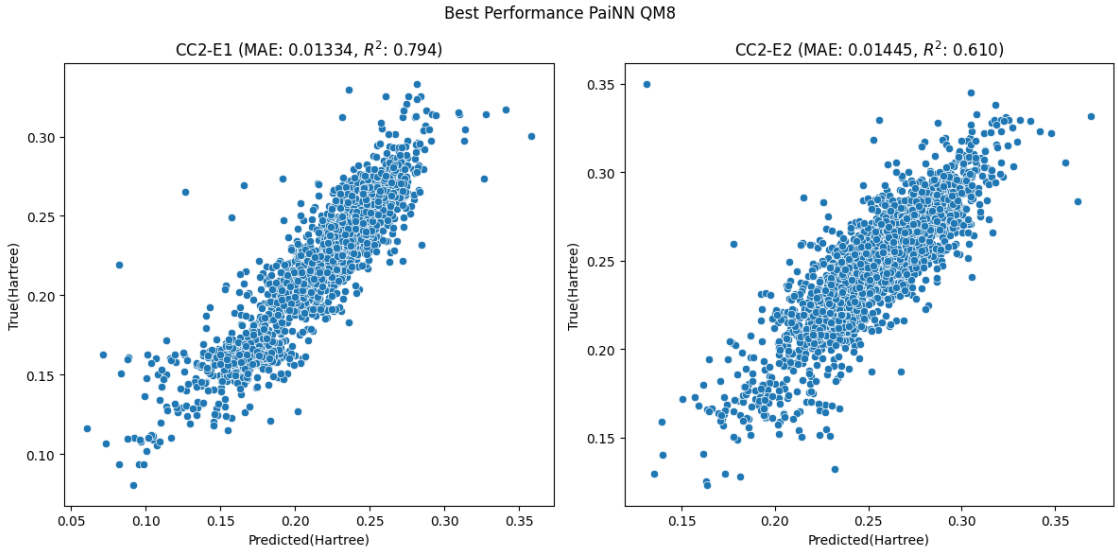

Figure S54: Parity Plot QM8 painn

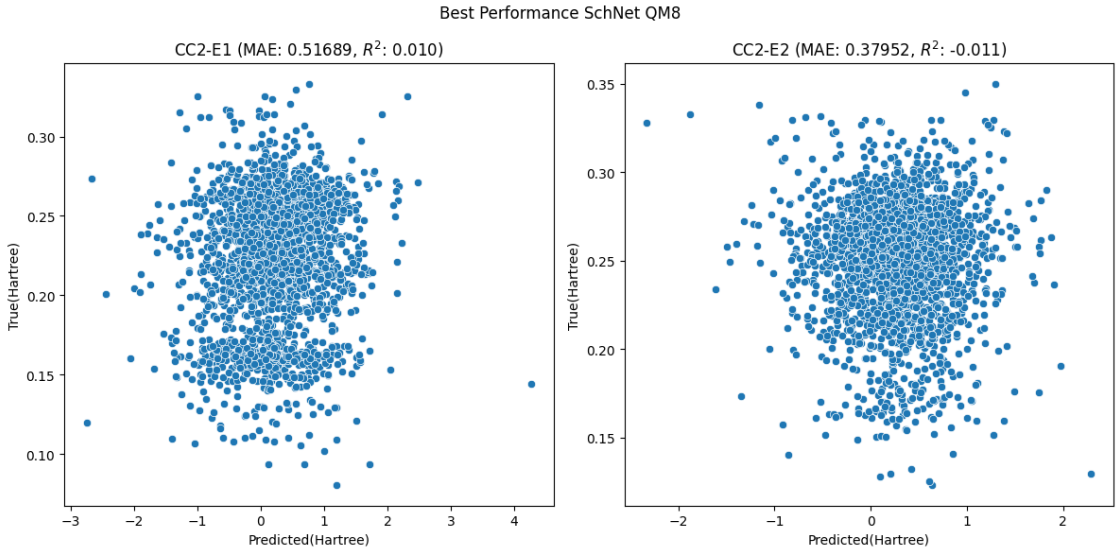

Figure S55: Parity Plot QM8 Schnet

References

(1) Heid, E.; Greenman, K. P.; Chung, Y.; Li, S.-C.; Graff, D. E.; Vermeire, F. H.; Wu, H.; Green, W. H.; McGill, C. J. Chemprop: A Machine Learning Package for Chemical Property Prediction. *Journal of Chemical Information and Modeling* **2023**, *64*, 9–17.

(2) Schütt, K.; Unke, O.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. Proceedings of the 38th International Conference on Machine Learning. 2021; pp 9377–9388.

(3) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature Communications* **2017**, *8*.

(4) Biewald, L. Experiment Tracking with Weights and Biases. 2020; `https://www.wandb.com/`, Software available from wandb.com.

TOC Graphic