Uncertainty Quantification for Molecular Property Predictions with Graph Neural Architecture Search

Shengli Jiang[1]*, Shiyi Qin[1], Reid C. Van Lehn[1],
Prasanna Balaprakash[2]and Victor M. Zavala[1,3]

[1]Department of Chemical and Biological Engineering
University of Wisconsin-Madison, 1415 Engineering Dr, Madison, WI 53706

[2]Computing and Computational Sciences Directorate
Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831

[3]Mathematics and Computer Science Division
Argonne National Laboratory, Lemont, IL 60439

S1 AutoGNNUQ

S1.1 Attention Functions

The attention functions used to calculate α_{vw} are shown in Table S1. For constant attention, α_{vw} is always 1. For GAT, α_{vw} is LeakyReLU($\mathbf{a}(\mathbf{W}\mathbf{h}_v||\mathbf{W}\mathbf{h}_w)$), where $\mathbf{a}$ s a trainable vector, $\mathbf{W}$ is a trainable weight matrix, $||$ denotes concatenation, and $\mathbf{h}_v$ and $\mathbf{h}_w$ are the node hidden feature vectors for nodes v and w, respectively. The SYM-GAT attention coefficient is procured by the summation of α_{vw} and α_{wv} from GAT. The COS attention type computes α_{vw} as $\mathbf{a}(\mathbf{W}\mathbf{h}_v||\mathbf{W}\mathbf{h}_w)$. In linear attention, α_{vw} is tanh($\mathbf{a}_l\mathbf{W}\mathbf{h}_v + \mathbf{W}\mathbf{h}_w$), with $\mathbf{a}l$ being a learnable vector. Generalized linear attention, or gen-linear attention, determines α_{vw} as $\mathbf{W}_G$tanh($\mathbf{W}\mathbf{h}_v + \mathbf{W}\mathbf{h}_w$), where $\mathbf{W}_G$ is a trainable matrix.

Table S1: Set of attention functions, where $||$ denotes the concatenation; $\mathbf{a}$, $\mathbf{a}_l$, and $\mathbf{a}_r$; and $\mathbf{W}_G$ is a trainable matrix.

Attention Mechanism	**Equations**
Constant	1
GAT	LeakyReLU($\mathbf{a}(\mathbf{W}\mathbf{h}_v\|\|\mathbf{W}\mathbf{h}_w)$)
SYM-GAT	$\alpha_{vw} + \alpha_{wv}$ based on GAT
COS	$\mathbf{a}(\mathbf{W}\mathbf{h}_v\|\|\mathbf{W}\mathbf{h}_w)$
Linear	tanh($\mathbf{a}_l\mathbf{W}\mathbf{h}_v + \mathbf{W}\mathbf{h}_w$)
Gen-Linear	$\mathbf{W}_G$tanh($\mathbf{W}\mathbf{h}_v + \mathbf{W}\mathbf{h}_w$)

*Corresponding Author: sjiang87@wisc.edu

S1.2 Architectures Discovered

Figures S1-S4 presents the optimal model architectures identified for the Lipo dataset using a random seed of 0, while Figures S5-S8 depicts the least effective models. In this context, input 0 corresponds to node features, input 1 to edge features, input 2 to edge pairs, and input 3 to node masks. A clear observation is that superior models generally incorporate more skip-connections and predominantly use the global attention pooling function. Conversely, the less effective models exhibit simpler architectures with rudimentary readout functions such as simple flatten.

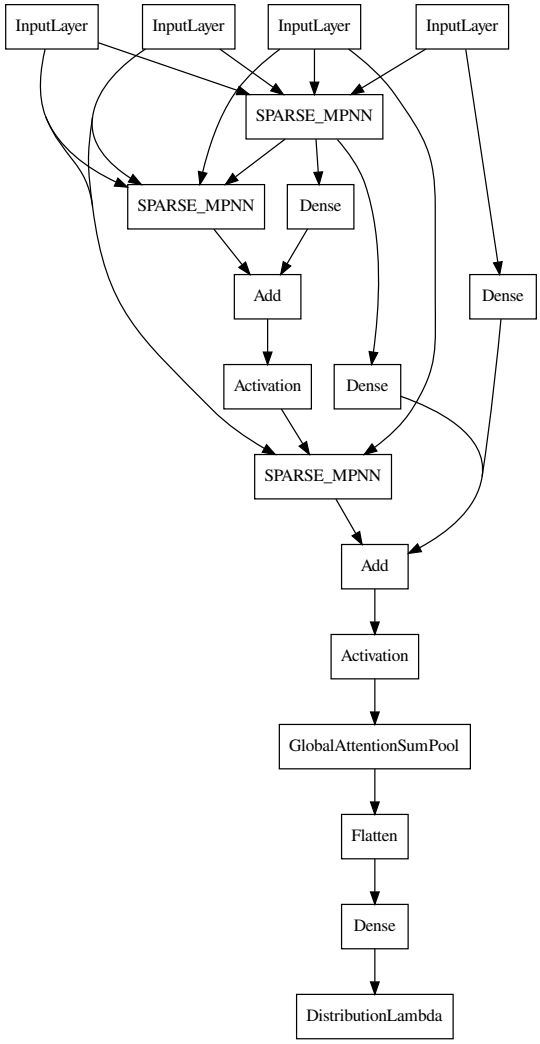

Figure S1: Model architectures demonstrating the highest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

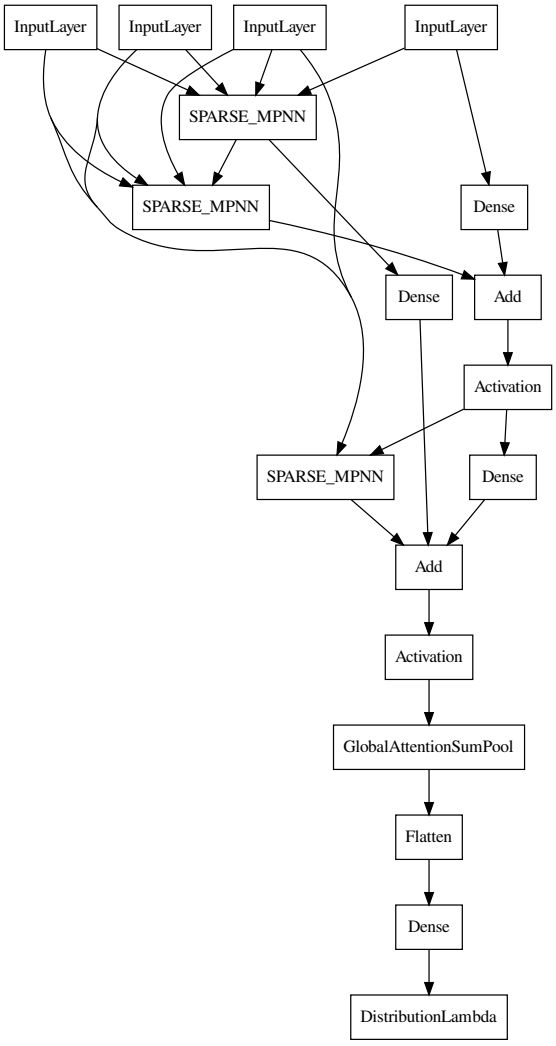

Figure S2: Model architectures demonstrating the second highest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

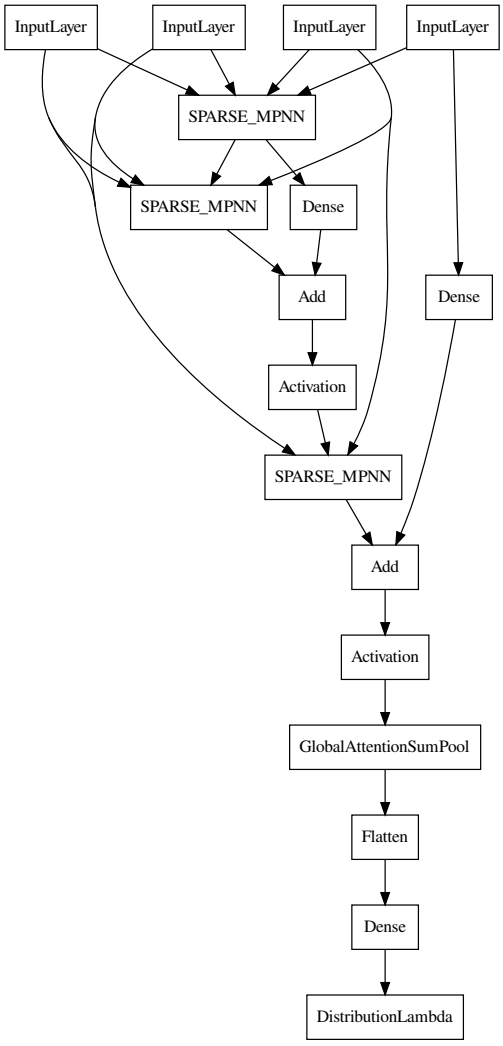

Figure S3: Model architectures demonstrating the third highest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

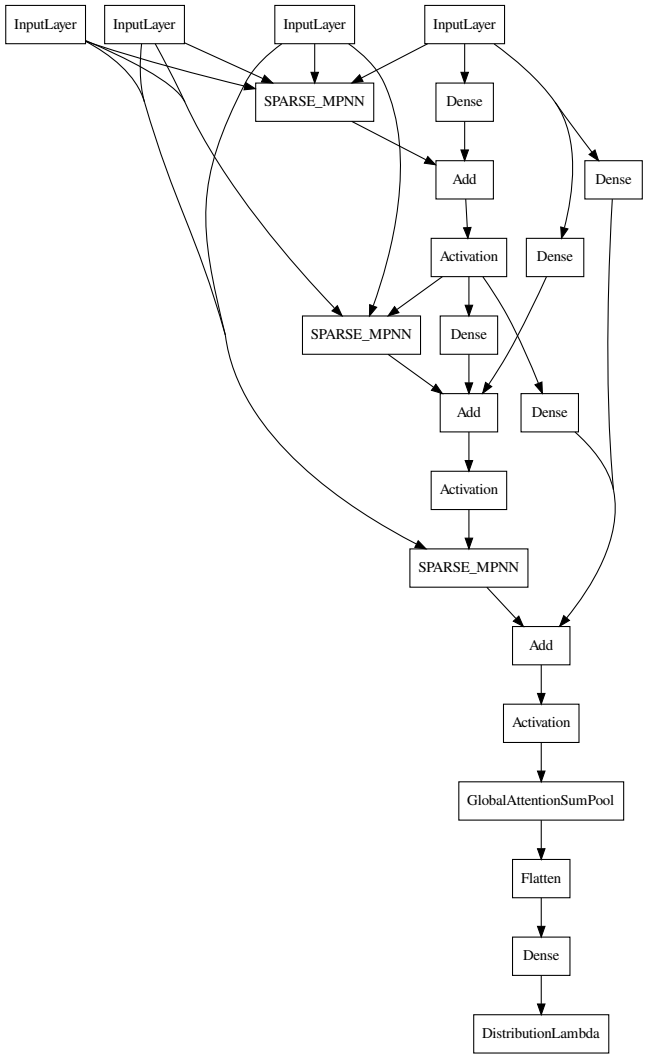

Figure S4: Model architectures demonstrating the fourth highest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

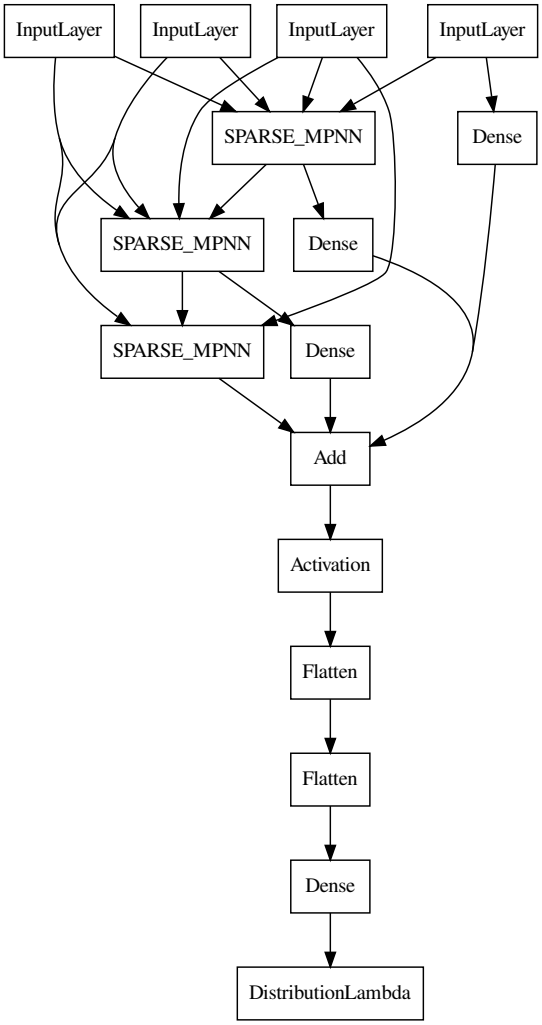

Figure S5: Model architectures demonstrating the lowest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

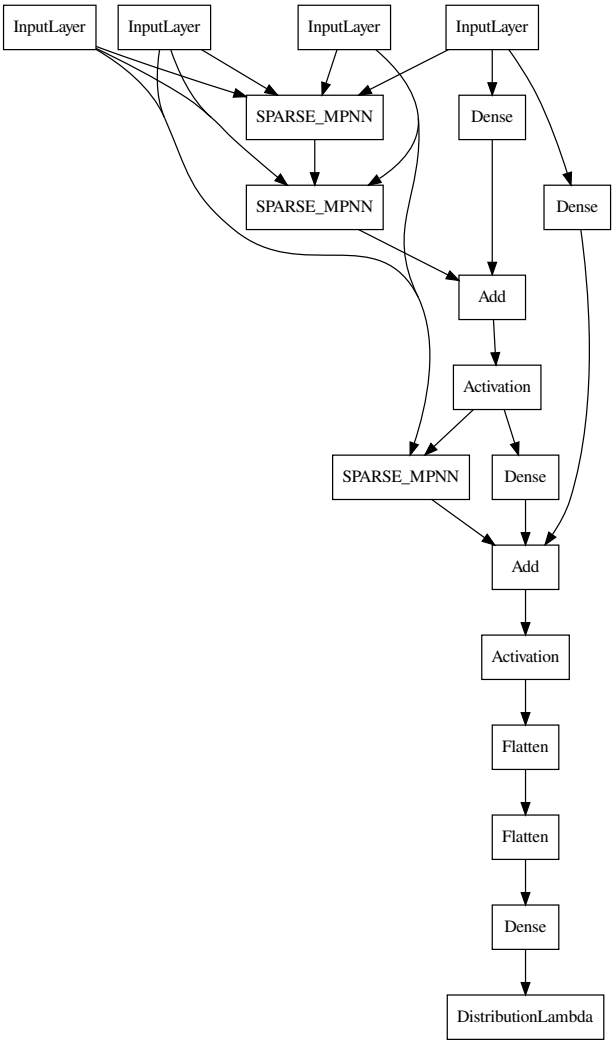

Figure S6: Model architectures demonstrating the second lowest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

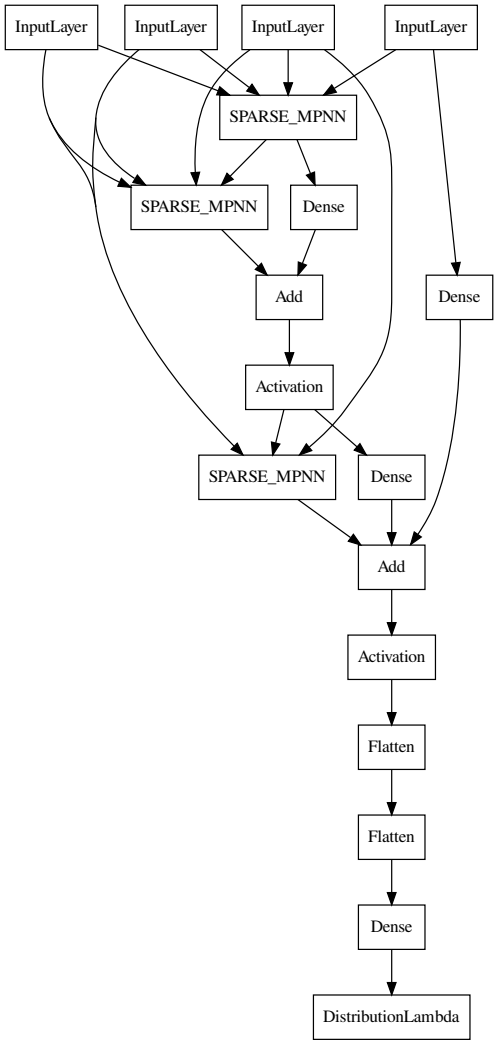

Figure S7: Model architectures demonstrating the third lowest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

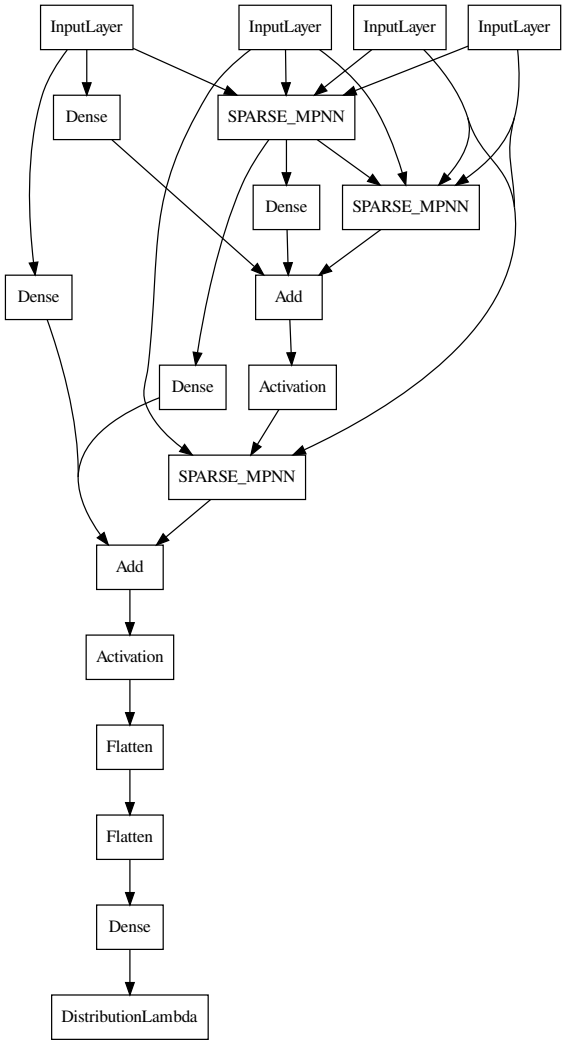

Figure S8: Model architectures demonstrating the fourth lowest search rewards (log-likelihood) for the Lipo dataset, using a random seed of 0.

S2 Prediction Parity Plots of AutoGNNUQ

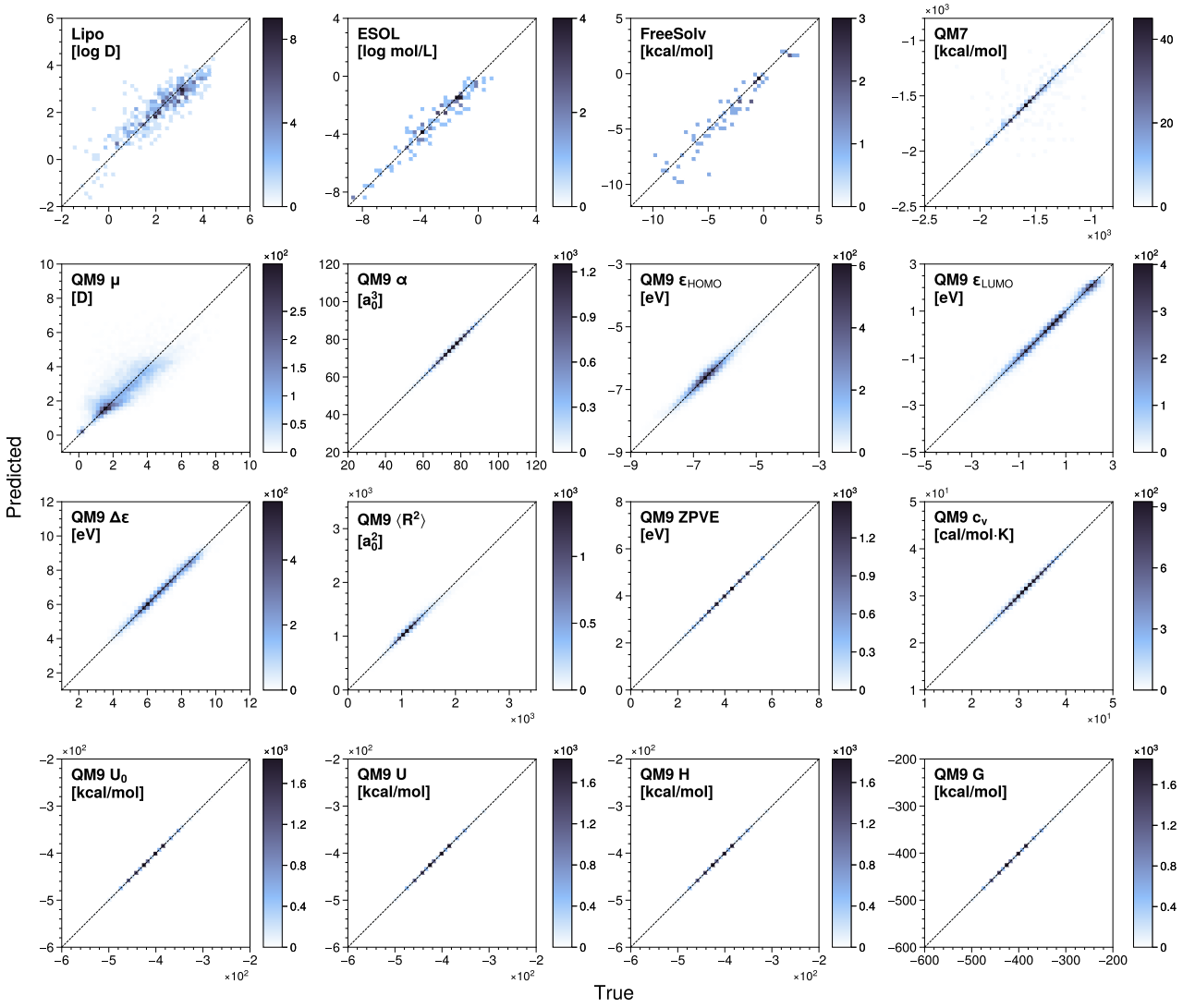

Figure S9: Prediction parity plot with a random seed of 1.

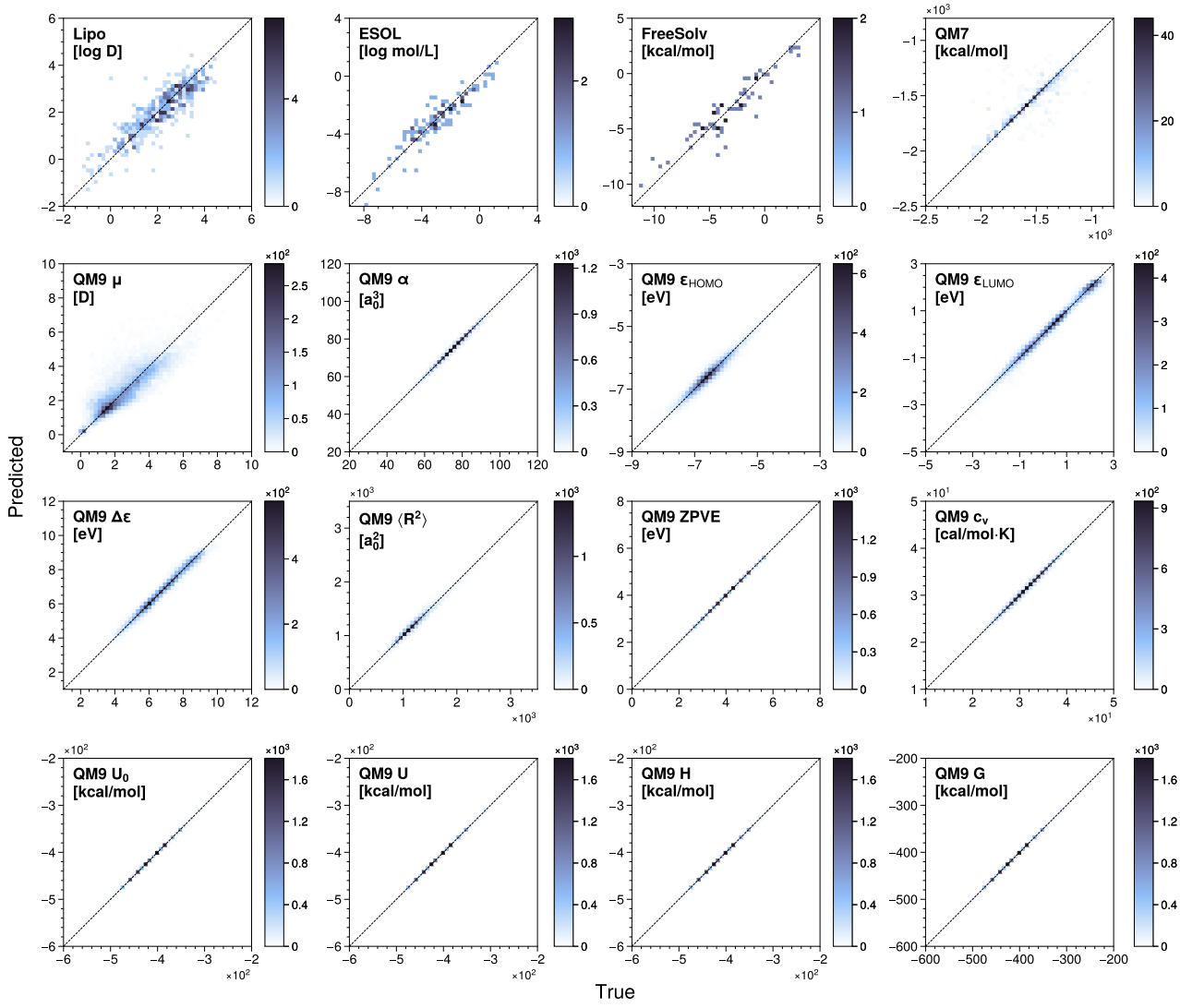

Figure S10: Prediction parity plot with a random seed of 2.

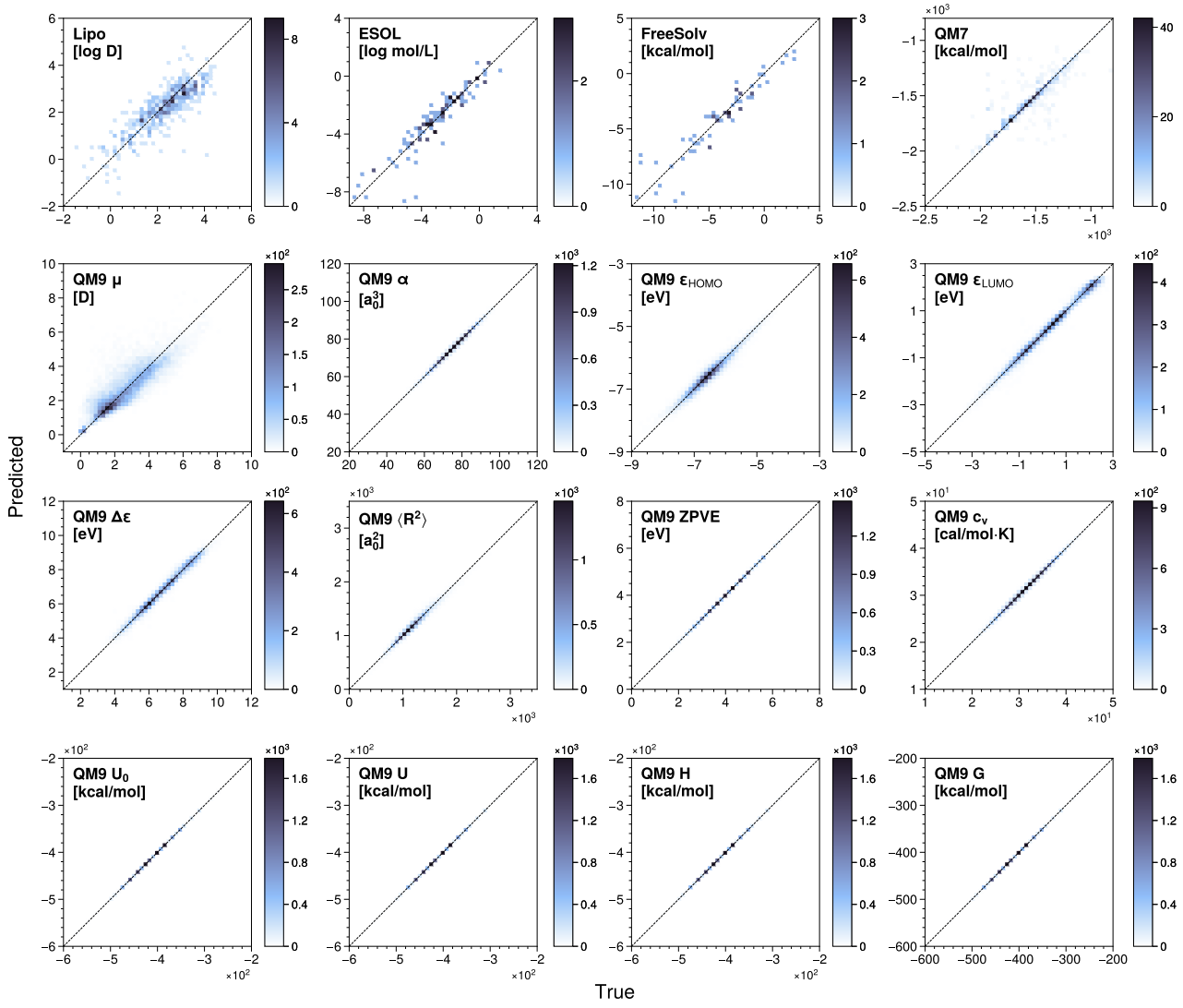

Figure S11: Prediction parity plot with a random seed of 3.

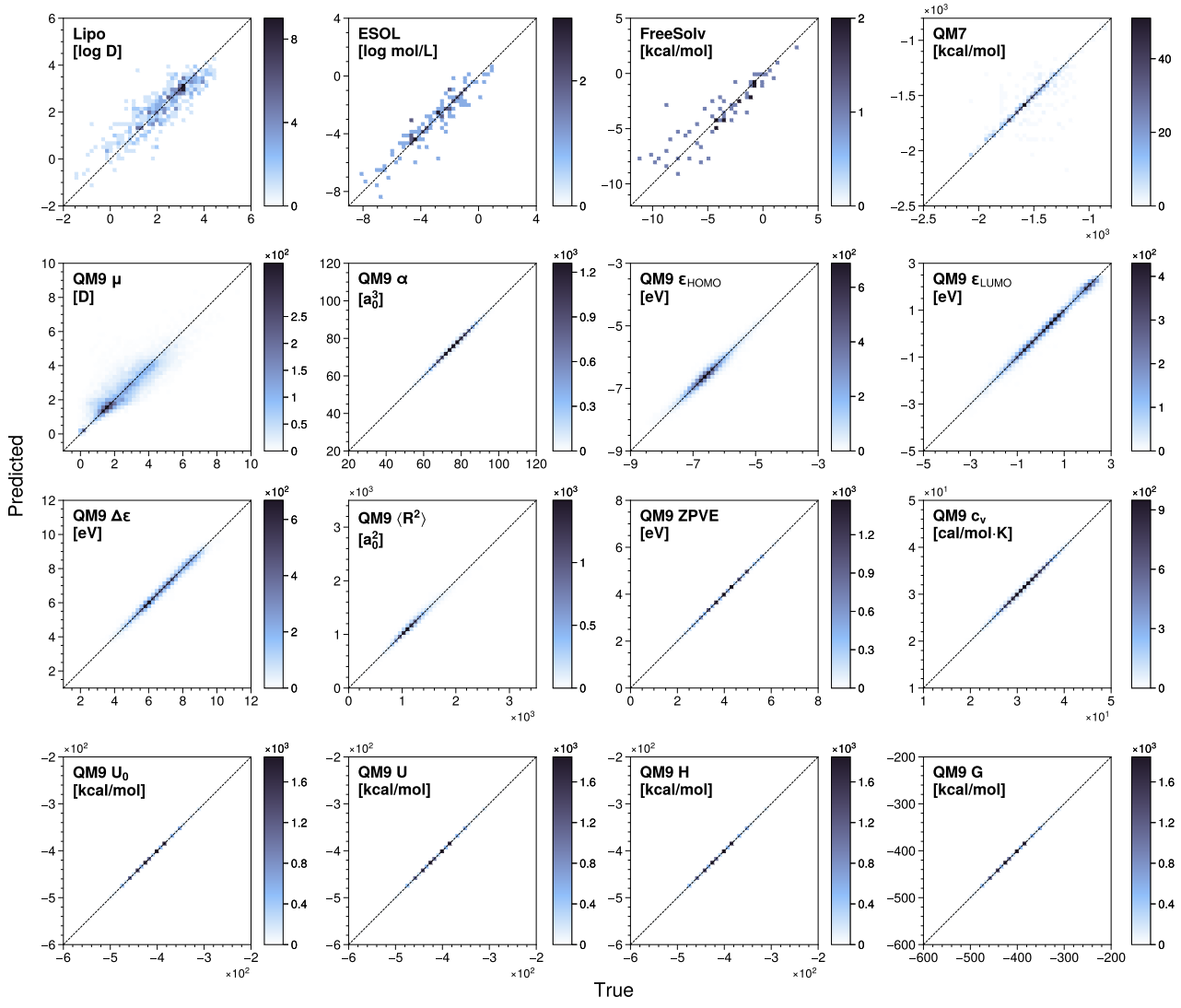

Figure S12: Prediction parity plot with a random seed of 4.

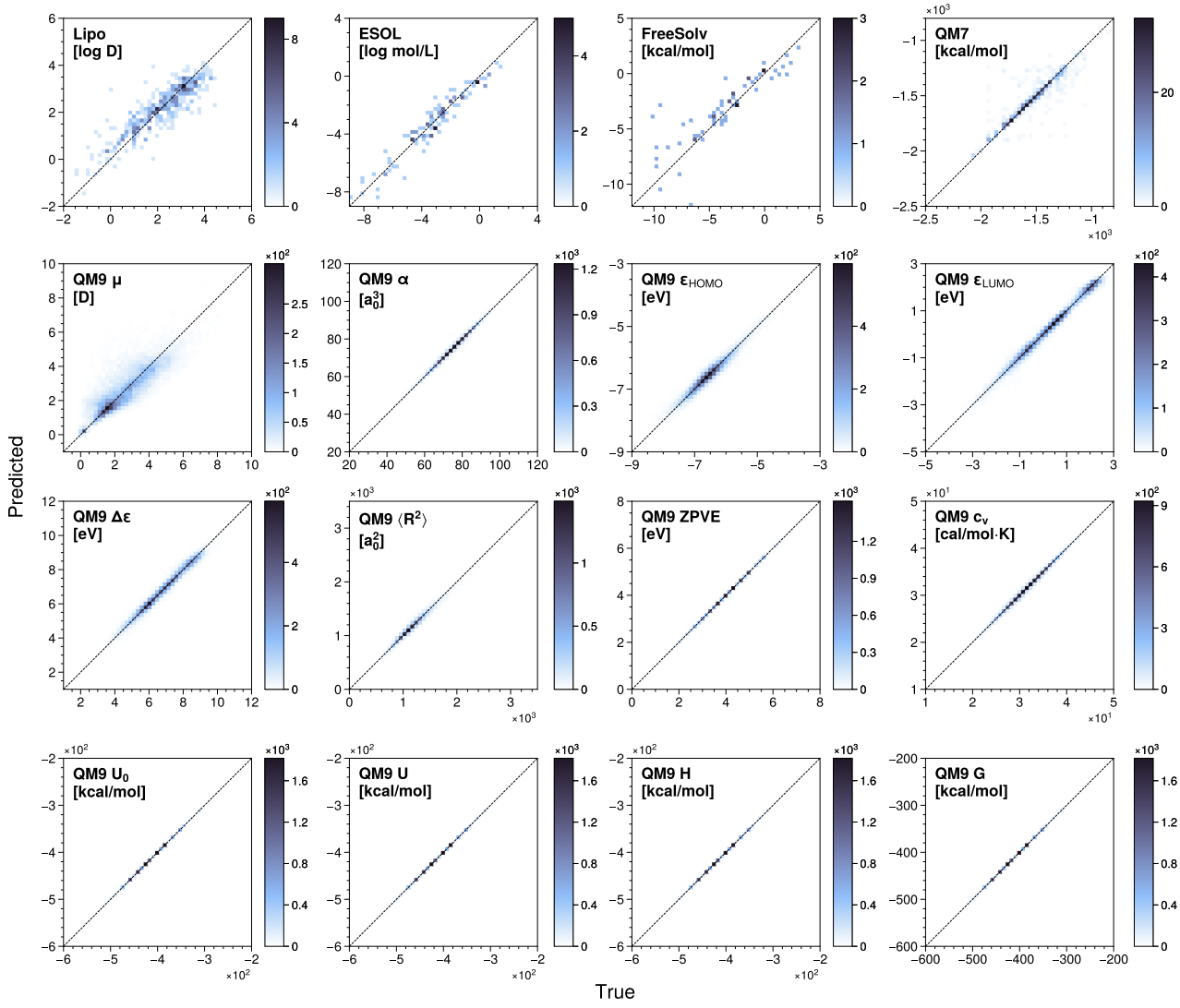

Figure S13: Prediction parity plot with a random seed of 5.

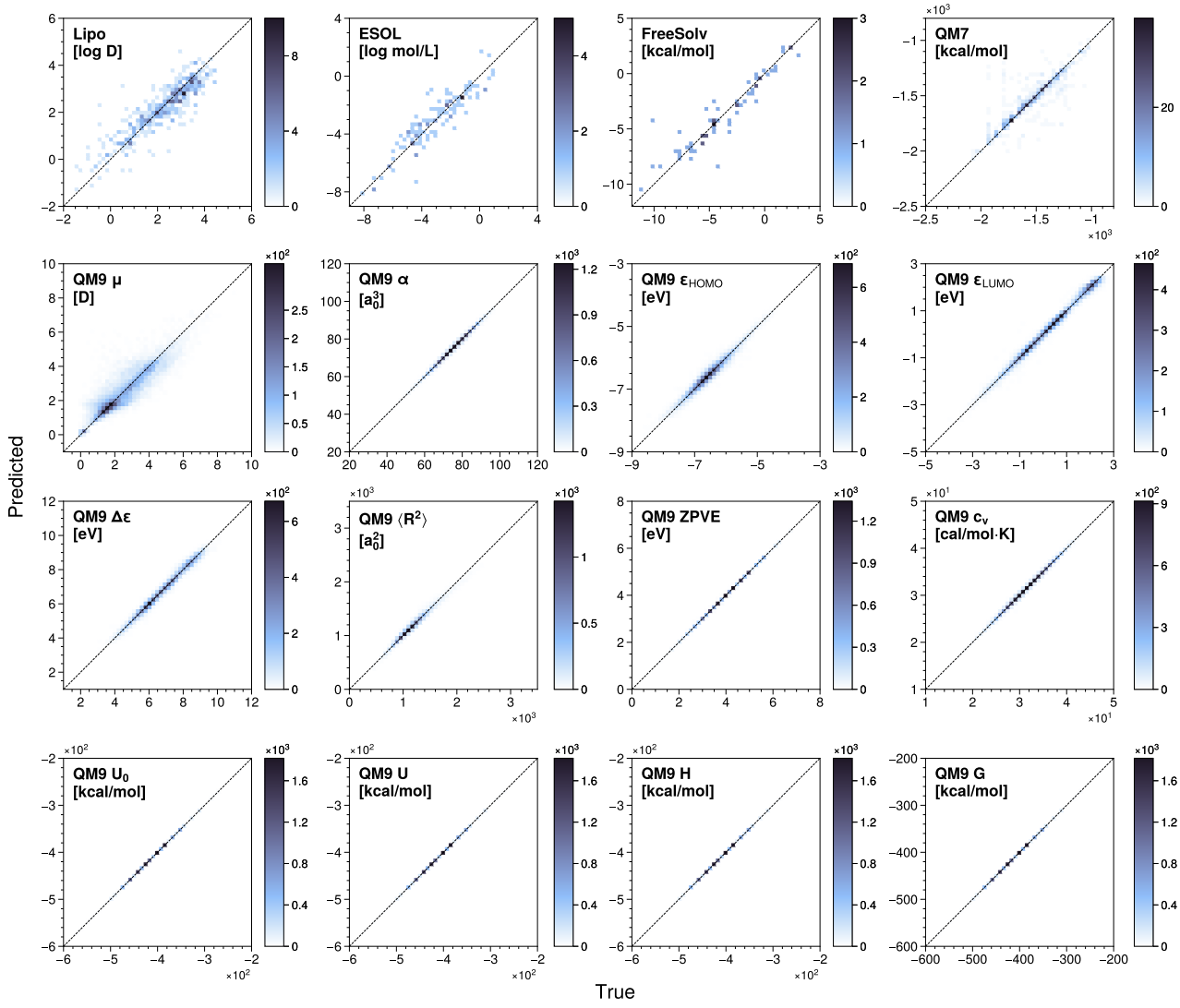

Figure S14: Prediction parity plot with a random seed of 6.

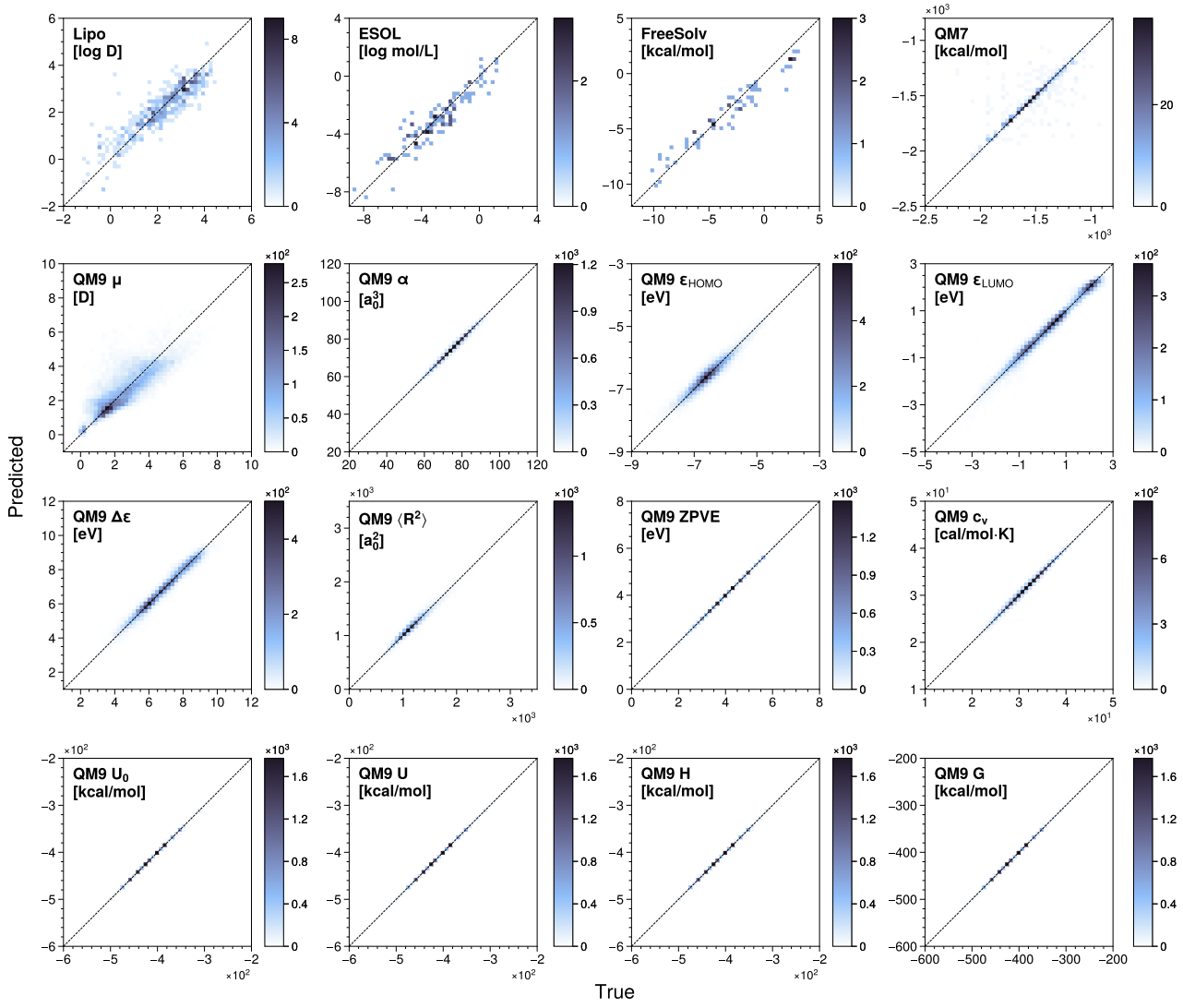

Figure S15: Prediction parity plot with a random seed of 7.

S3 Prediction Error and Uncertainty of AutoGNNUQ

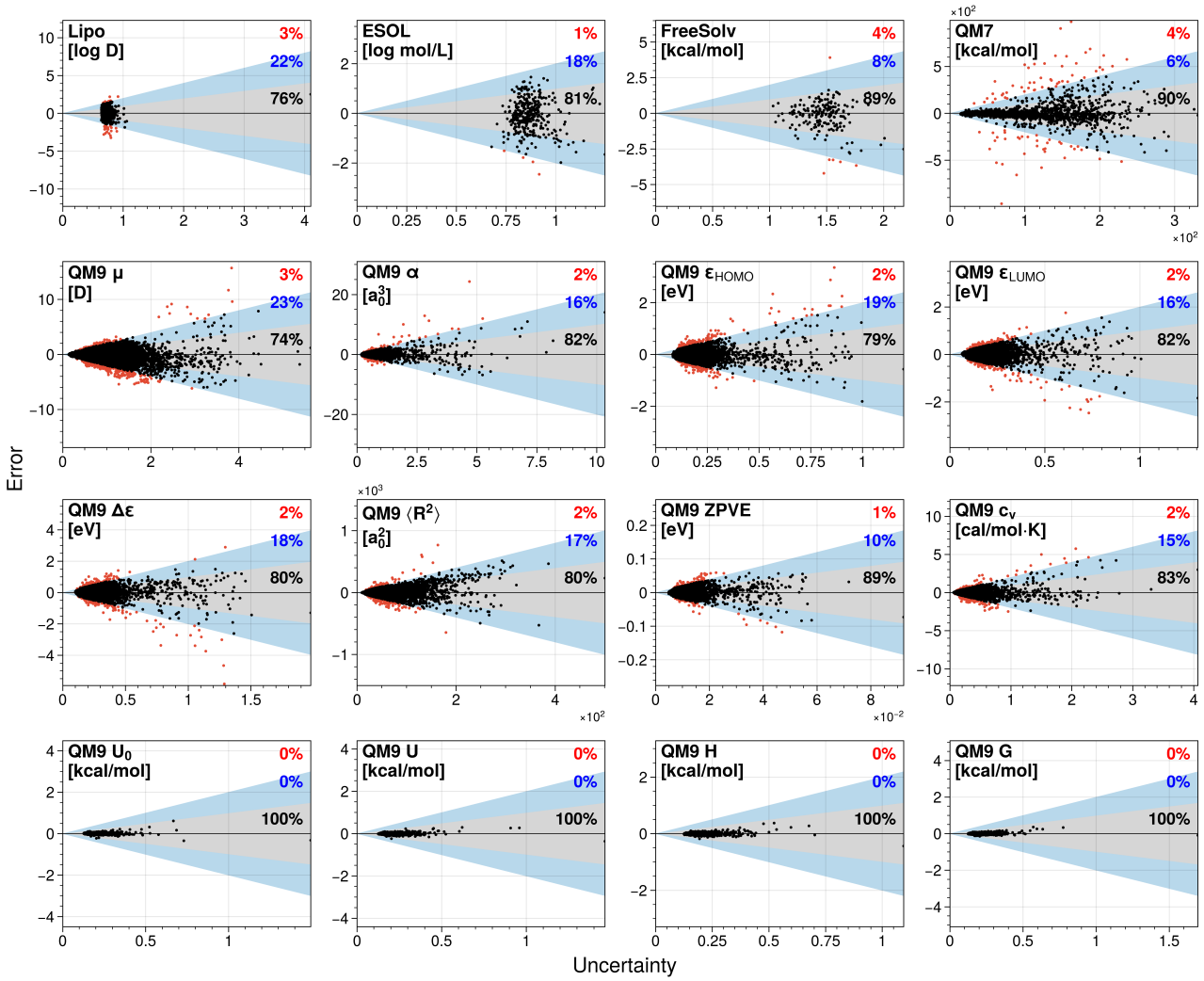

Figure S16: Prediction error vs uncertainty with a random seed of 1.

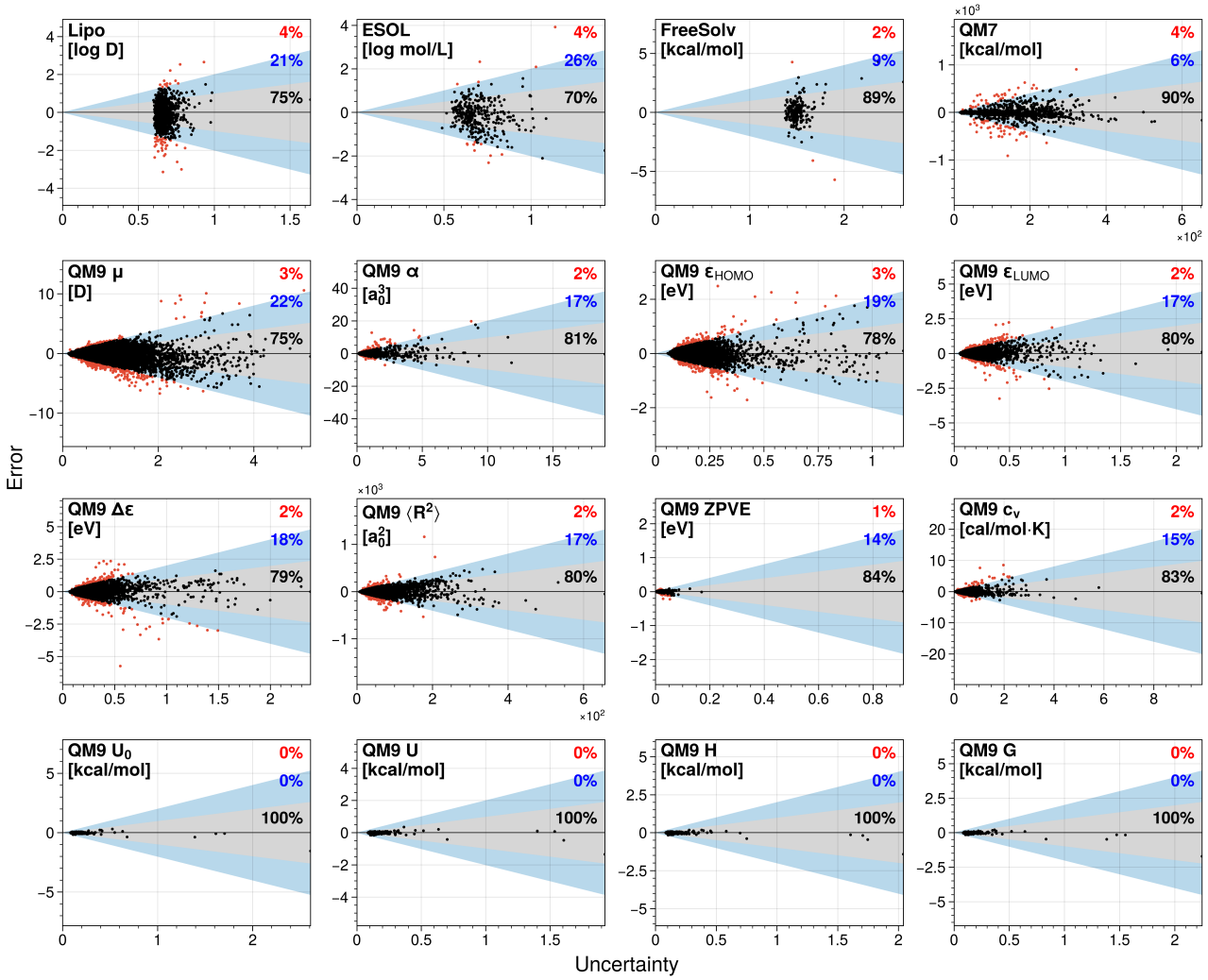

Figure S17: Prediction error vs uncertainty with a random seed of 2.

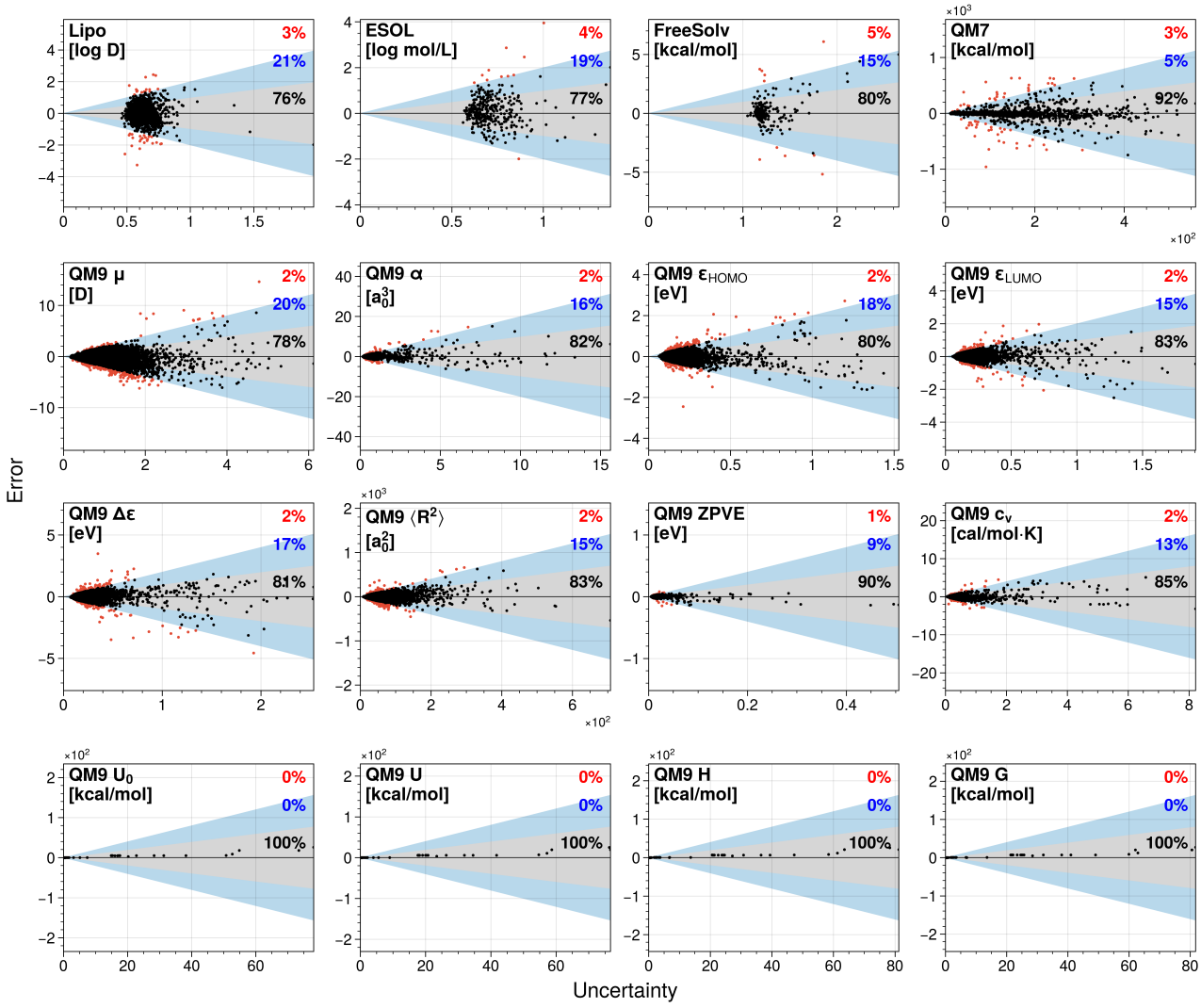

Figure S18: Prediction error vs uncertainty with a random seed of 3.

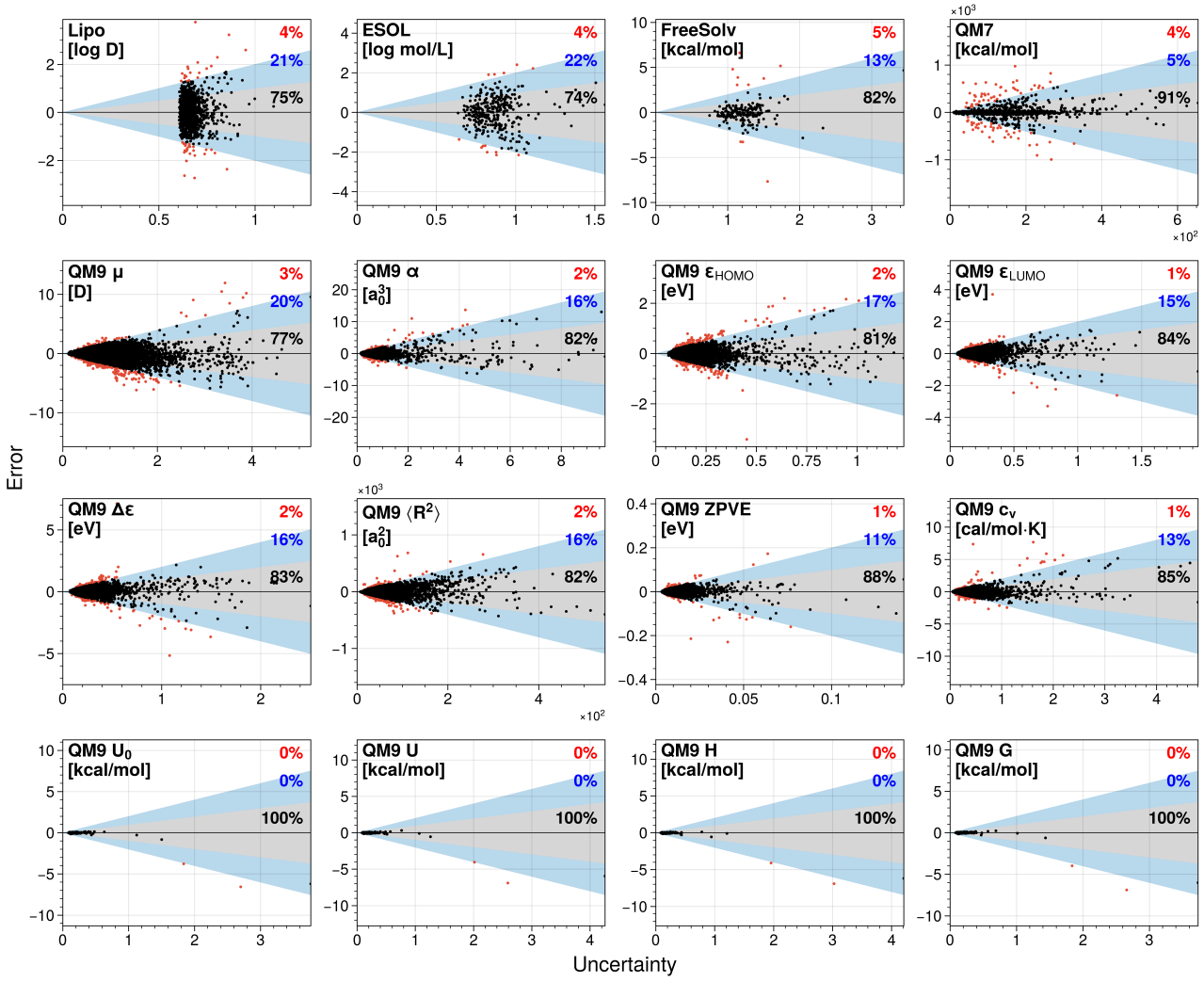

Figure S19: Prediction error vs uncertainty with a random seed of 4.

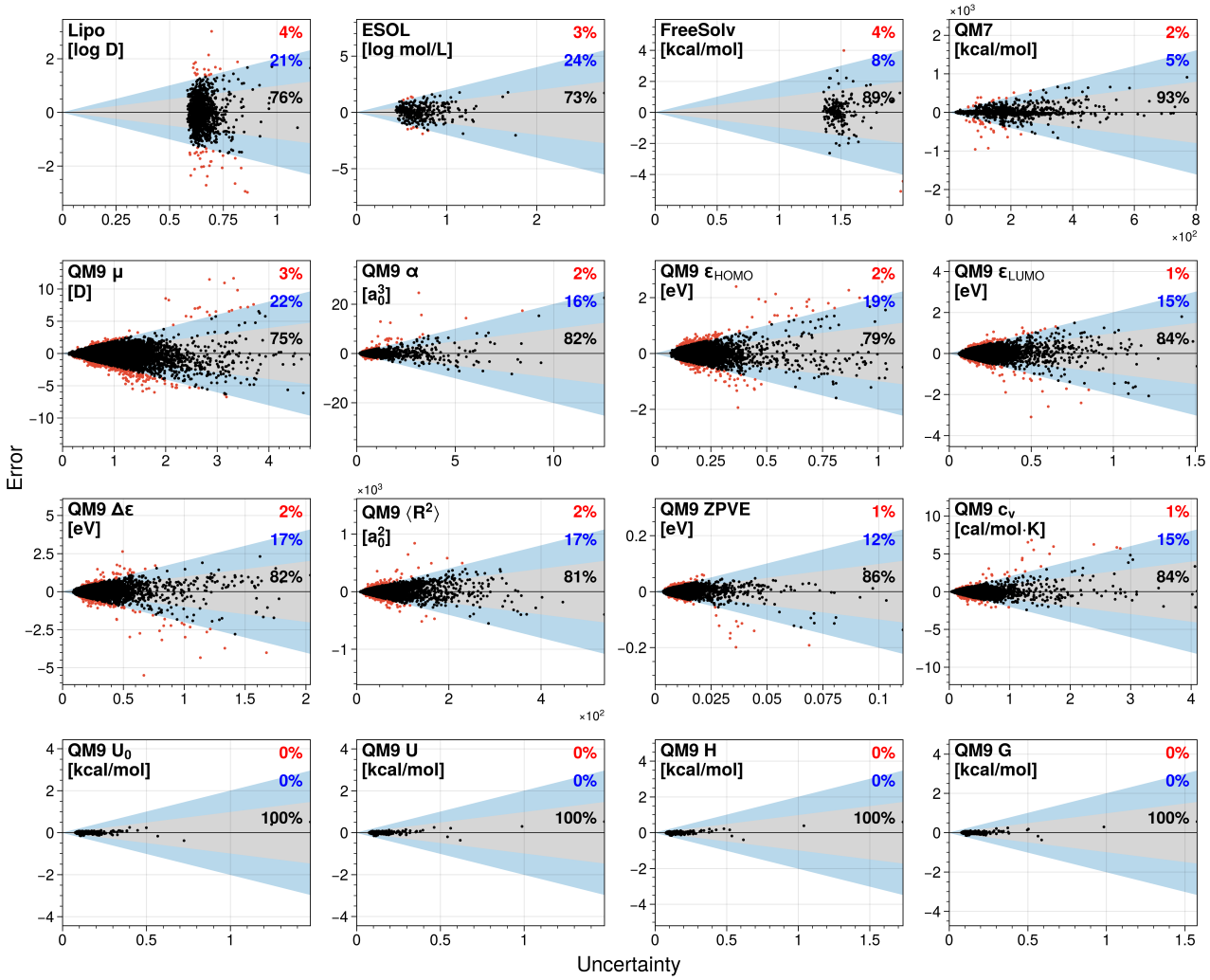

Figure S20: Prediction error vs uncertainty with a random seed of 5.

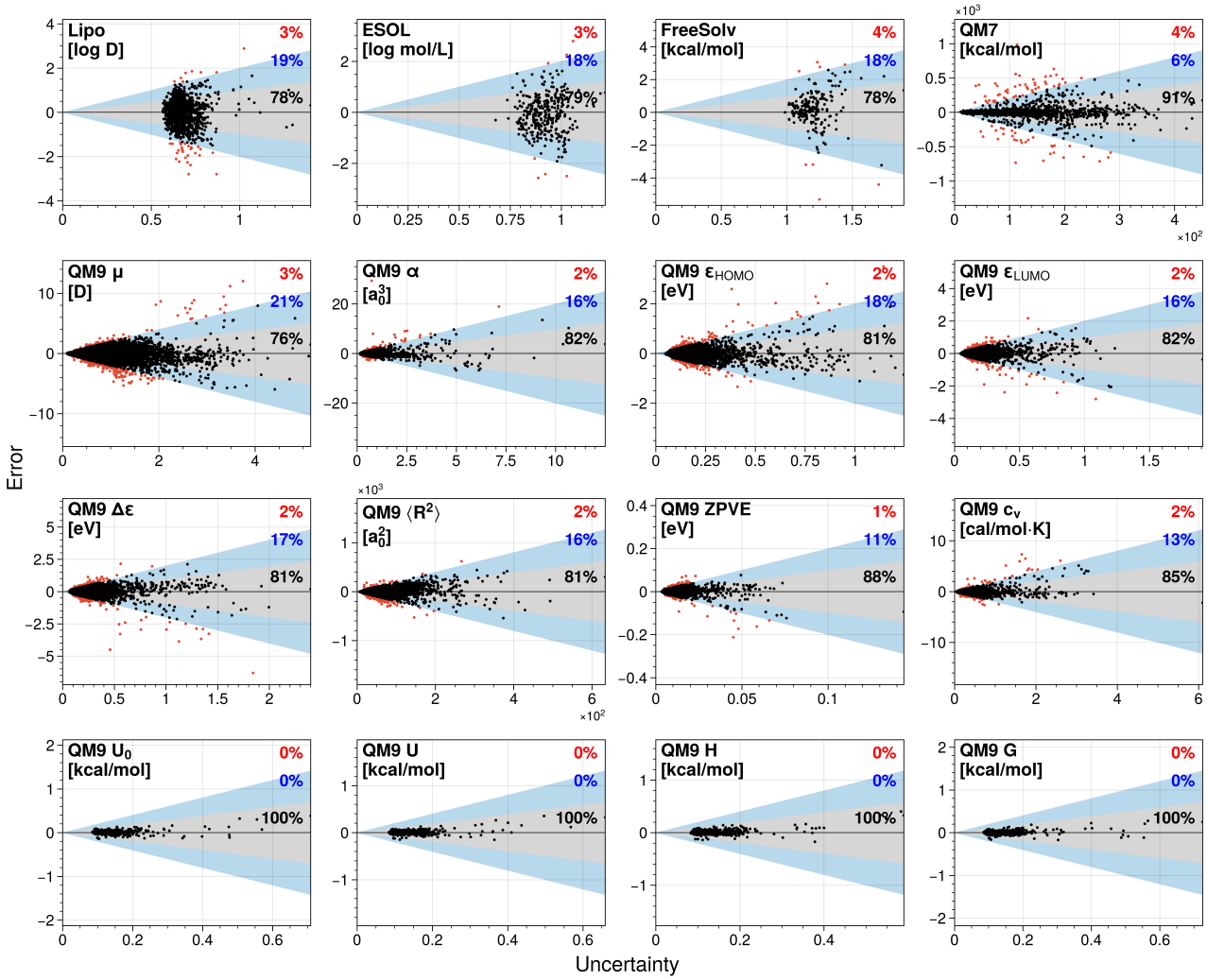

Figure S21: Prediction error vs uncertainty with a random seed of 6.

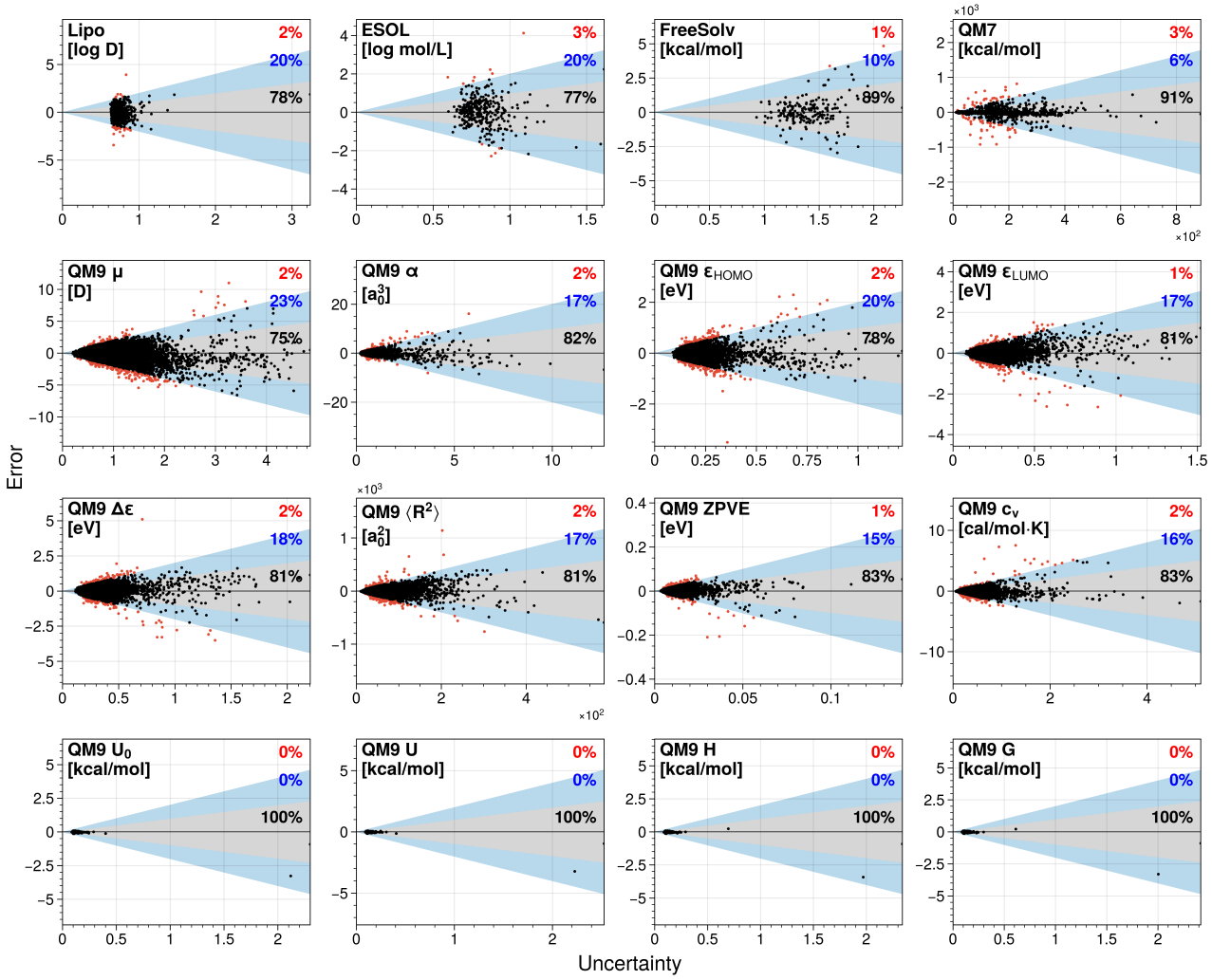

Figure S22: Prediction error vs uncertainty with a random seed of 7.

S4 TSNE of AutoGNNUQ

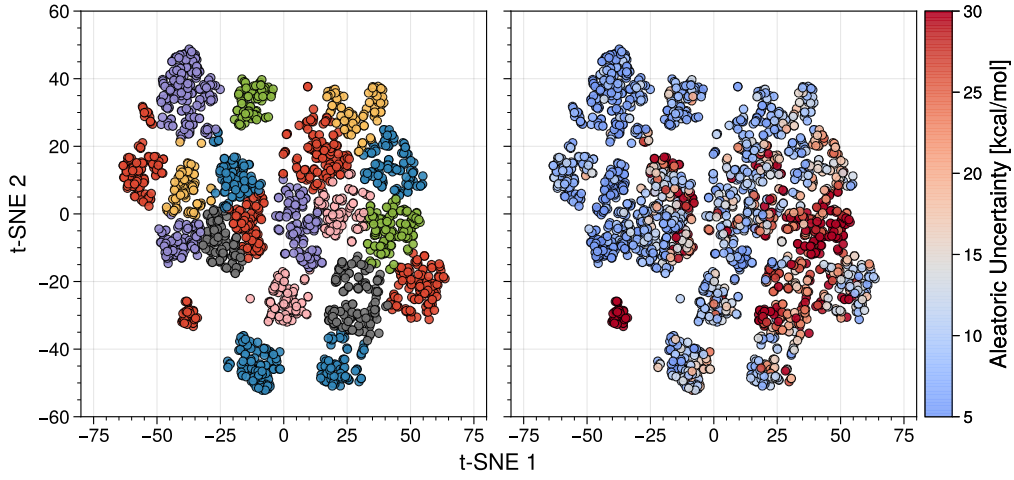

Figure S23: t-SNE plot of QM7 MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Aleatoric uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

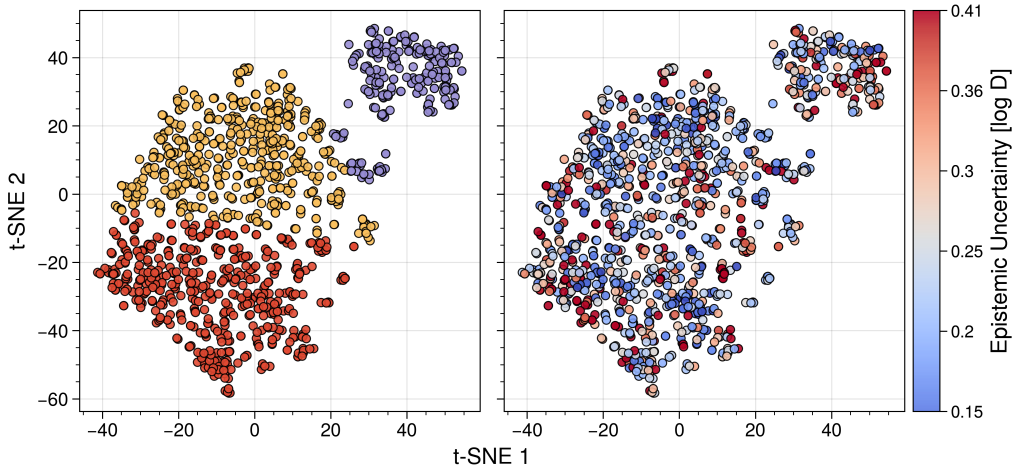

Figure S24: t-SNE plot of Lipo MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Epistemic uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

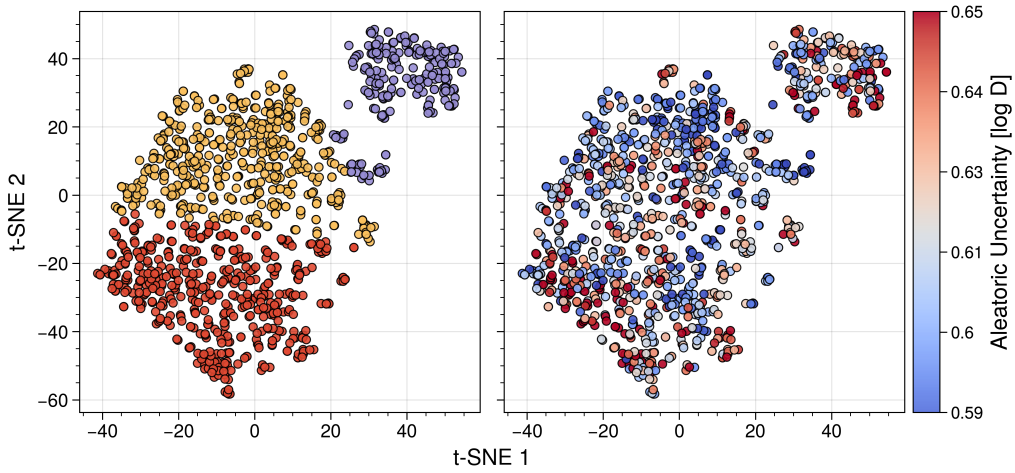

Figure S25: t-SNE plot of Lipo MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Aleatoric uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

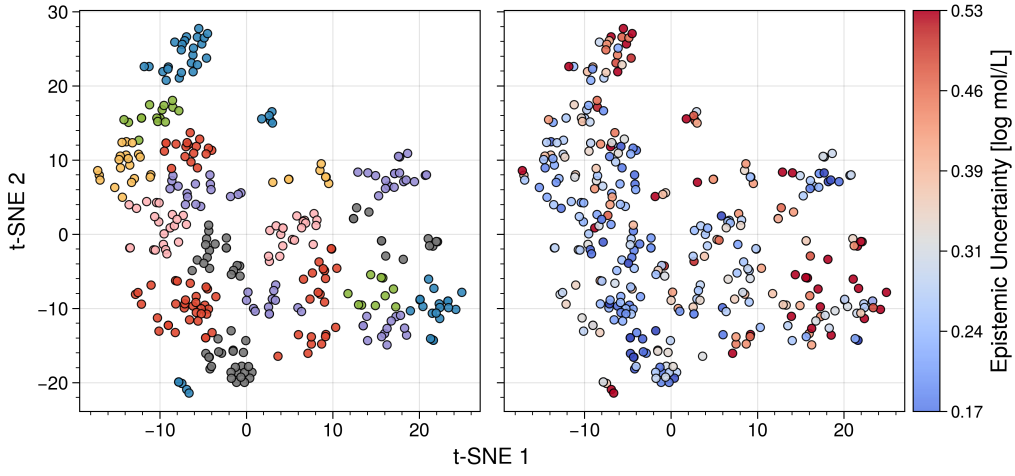

Figure S26: t-SNE plot of ESOL MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Epistemic uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

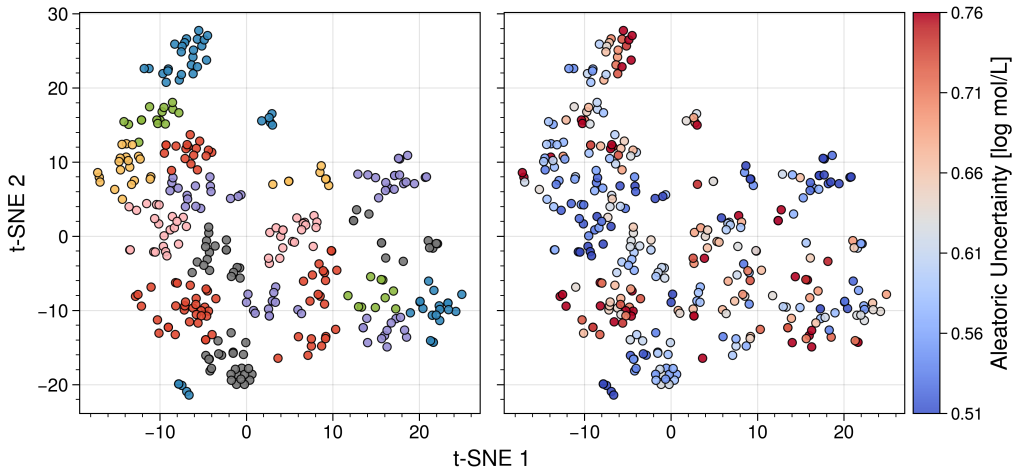

Figure S27: t-SNE plot of ESOL MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Aleatoric uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

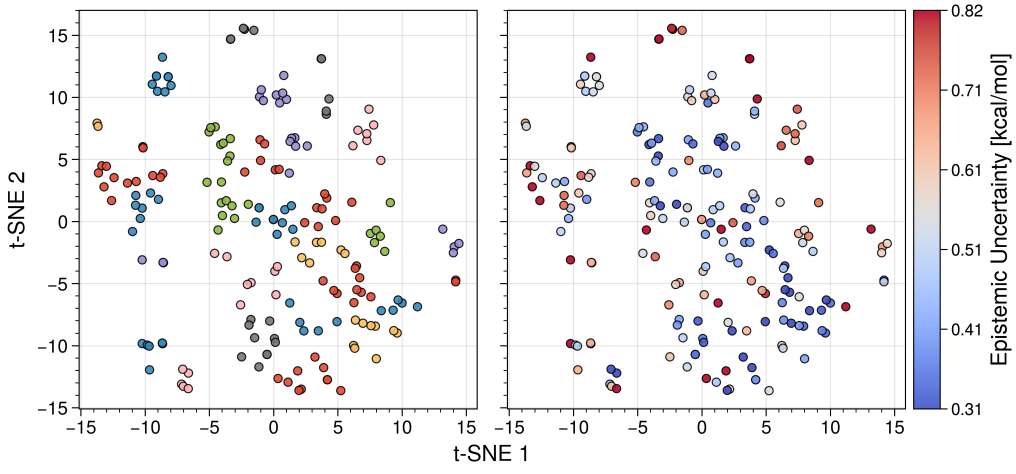

Figure S28: t-SNE plot of FreeSolv MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Epistemic uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

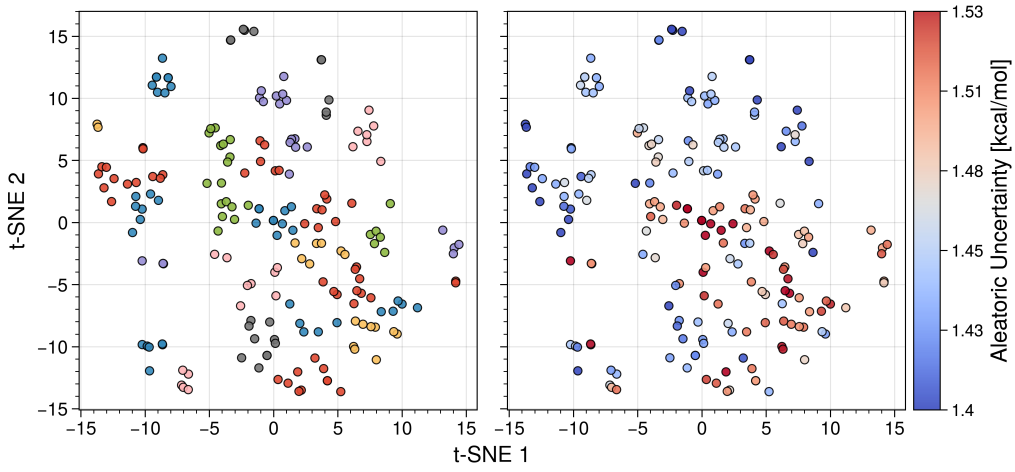

Figure S29: t-SNE plot of FreeSolv MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Aleatoric uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

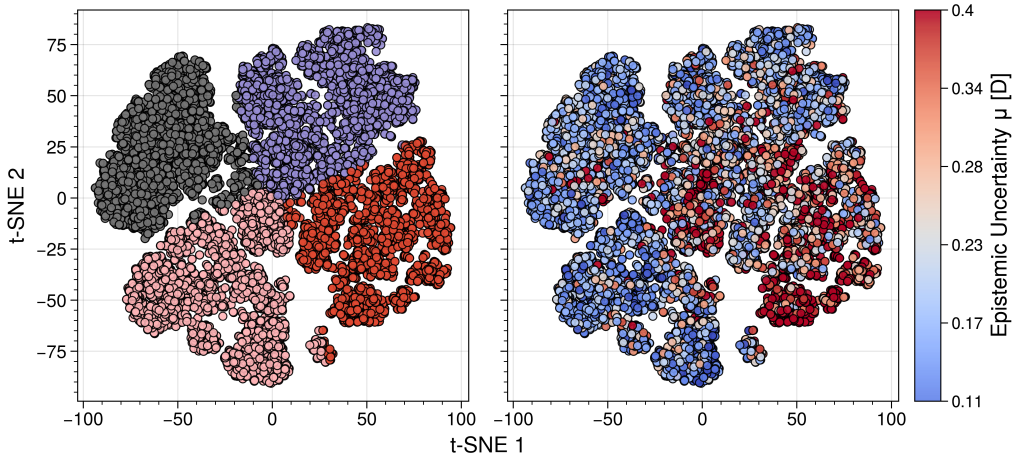

Figure S30: t-SNE plot of QM9 μ MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Epistemic uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

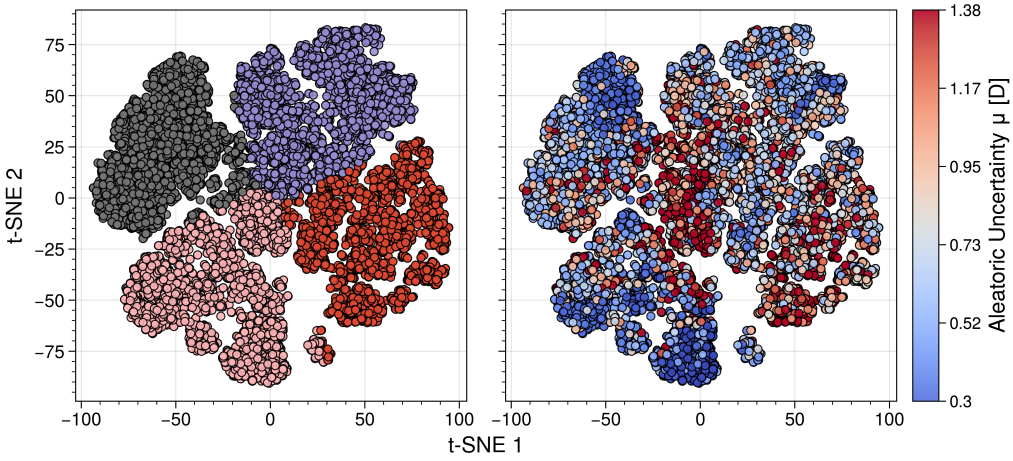

Figure S31: t-SNE plot of QM9 μ MACCS keys for random seed 0. (a) Clustering by k-means algorithm. (b) Aleatoric uncertainties associated with each molecule. Low uncertainties are represented in blue and high uncertainties in red.

S5 Confidence-Based Recalibration

Table S2: Confidence-based calibration ratios. Mean value reported with standard deviation in parentheses.

Dataset	Property	AutoGNNUQ	AutoGNNUQ-Simple	MC Dropout	Random Ensemble
Lipo	logP	0.846 (0.035)	1.003 (0.087)	0.973 (0.080)	0.827 (0.028)
ESOL	S_{water}	0.844 (0.044)	0.894 (0.030)	1.003 (0.059)	0.652 (0.063)
FreeSolv	ΔG_{hyd}	0.562 (0.067)	0.738 (0.131)	0.770 (0.150)	0.522 (0.050)
QM7	$\Delta H^{\circ}_{\mathrm{atom}}$	0.276 (0.049)	0.252 (0.025)	0.380 (0.080)	0.358 (0.052)
QM9	μ	0.847 (0.022)	0.919 (0.070)	0.874 (0.030)	0.799 (0.037)
	α	0.721 (0.015)	0.730 (0.038)	0.507 (0.092)	0.441 (0.044)
	ϵ_{HOMO}	0.773 (0.020)	0.792 (0.050)	0.768 (0.021)	0.634 (0.023)
	ϵ_{LUMO}	0.709 (0.020)	0.726 (0.045)	0.669 (0.037)	0.546 (0.041)
	$\Delta\epsilon$	0.734 (0.020)	0.738 (0.050)	0.706 (0.024)	0.581 (0.039)
	$\langle R^2 \rangle$	0.732 (0.019)	0.751 (0.044)	0.646 (0.050)	0.505 (0.038)
	ZPVE	0.617 (0.048)	0.653 (0.069	0.514 (0.115)	0.429 (0.082)
	c_v	0.670 (0.023)	0.695 (0.055)	0.566 (0.033)	0.465 (0.061)
	U_0	0.126 (0.035)	0.117 (0.022)	0.305 (0.046)	0.257 (0.063)
	U	0.112 (0.013)	0.116 (0.024)	0.305 (0.046)	0.255 (0.067)
	H	0.134 (0.022)	0.107 (0.013)	0.305 (0.046)	0.255 (0.067)
	G	0.114 (0.020)	0.124 (0.023)	0.305 (0.046)	0.256 (0.064)

Table S3: Confidence-based calibration metrics of random ensemble before and after recalibration based on the validation dataset. Mean value reported with standard deviation in parentheses.

Dataset	Property	ECE		MCE		MCA	
		Before	After	Before	After	Before	After
Lipo	logP	0.063 (0.008)	**0.009 (0.005)**	0.106 (0.013)	**0.023 (0.008)**	0.063 (0.008)	**0.009 (0.005)**
ESOL	S_{water}	0.135 (0.018)	**0.021 (0.013)**	0.223 (0.032)	**0.046 (0.019)**	0.136 (0.018)	**0.022 (0.013)**
FreeSolv	ΔG_{hyd}	0.185 (0.015)	**0.035 (0.018)**	0.317 (0.037)	**0.079 (0.029)**	0.187 (0.015)	**0.035 (0.018)**
QM7	$\Delta H^{\circ}_{\text{atom}}$	0.241 (0.022)	**0.020 (0.003)**	0.442 (0.041)	**0.081 (0.014)**	0.243 (0.022)	**0.020 (0.003)**
QM9	μ	0.072 (0.012)	**0.006 (0.001)**	0.114 (0.022)	**0.016 (0.003)**	0.073 (0.012)	**0.006 (0.001)**
	α	0.242 (0.021)	**0.022 (0.003)**	0.435 (0.040)	**0.063 (0.011)**	0.245 (0.021)	**0.023 (0.003)**
	ϵ_{HOMO}	0.141 (0.011)	**0.005 (0.001)**	0.224 (0.018)	**0.014 (0.003)**	0.142 (0.011)	**0.005 (0.001)**
	ϵ_{LUMO}	0.194 (0.017)	**0.019 (0.005)**	0.345 (0.032)	**0.063 (0.021)**	0.196 (0.018)	**0.019 (0.005)**
	$\Delta\epsilon$	0.168 (0.020)	**0.017 (0.006)**	0.290 (0.032)	**0.038 (0.012)**	0.169 (0.020)	**0.017 (0.006)**
	$\langle R^2 \rangle$	0.203 (0.017)	**0.018 (0.007)**	0.356 (0.024)	**0.039 (0.015)**	0.205 (0.017)	**0.019 (0.007)**
	ZPVE	0.263 (0.043)	**0.044 (0.008)**	0.525 (0.086)	**0.138 (0.020)**	0.266 (0.043)	**0.044 (0.008)**
	c_v	0.237 (0.027)	**0.027 (0.007)**	0.438 (0.046)	**0.084 (0.025)**	0.240 (0.028)	**0.028 (0.007)**
	U_0	0.344 (0.031)	**0.042 (0.014)**	0.656 (0.043)	**0.106 (0.023)**	0.348 (0.032)	**0.043 (0.014)**
	U	0.345 (0.033)	**0.044 (0.010)**	0.656 (0.043)	**0.106 (0.024)**	0.349 (0.033)	**0.044 (0.010)**
	H	0.345 (0.033)	**0.044 (0.010)**	0.656 (0.044)	**0.107 (0.022)**	0.349 (0.033)	**0.044 (0.011)**
	G	0.345 (0.032)	**0.044 (0.011)**	0.657 (0.044)	**0.105 (0.025)**	0.348 (0.032)	**0.044 (0.011)**

Table S4: Confidence-based calibration metrics of MC dropout before and after recalibration based on the validation dataset. Mean value reported with standard deviation in parentheses.

Dataset	Property	ECE		MCE		MCA	
		Before	After	Before	After	Before	After
Lipo	logP	0.020 (0.010)	**0.014 (0.008)**	0.043 (0.016)	**0.031 (0.011)**	0.020 (0.010)	**0.014 (0.008)**
ESOL	S_{water}	0.017 (0.005)	**0.019 (0.005)**	0.043 (0.011)	**0.048 (0.009)**	0.017 (0.005)	**0.020 (0.005)**
FreeSolv	ΔG_{hyd}	0.090 (0.049)	**0.035 (0.009)**	0.174 (0.080)	**0.094 (0.026)**	0.091 (0.049)	**0.035 (0.009)**
QM7	$\Delta H^{\circ}_{\text{atom}}$	0.215 (0.030)	**0.027 (0.007)**	0.383 (0.059)	**0.110 (0.022)**	0.217 (0.030)	**0.028 (0.007)**
QM9	μ	0.042 (0.009)	**0.005 (0.002)**	0.067 (0.014)	**0.012 (0.006)**	0.043 (0.009)	**0.005 (0.002)**
	α	0.200 (0.040)	**0.004 (0.001)**	0.322 (0.068)	**0.010 (0.002)**	0.202 (0.041)	**0.004 (0.001)**
	ϵ_{HOMO}	0.081 (0.007)	**0.006 (0.002)**	0.128 (0.011)	**0.016 (0.004)**	0.082 (0.007)	**0.006 (0.002)**
	ϵ_{LUMO}	0.120 (0.012)	**0.006 (0.001)**	0.189 (0.019)	**0.018 (0.005)**	0.121 (0.012)	**0.006 (0.001)**
	$\Delta\epsilon$	0.103 (0.011)	**0.006 (0.001)**	0.162 (0.016)	**0.018 (0.003)**	0.104 (0.011)	**0.006 (0.001)**
	$\langle R^2 \rangle$	0.130 (0.021)	**0.006 (0.002)**	0.207 (0.033)	**0.016 (0.004)**	0.131 (0.021)	**0.006 (0.002)**
	ZPVE	0.191 (0.057)	**0.008 (0.005)**	0.306 (0.095)	**0.023 (0.013)**	0.193 (0.058)	**0.008 (0.005)**
	c_v	0.166 (0.013)	**0.007 (0.001)**	0.262 (0.023)	**0.019 (0.004)**	0.168 (0.014)	**0.007 (0.002)**
	U_0	0.299 (0.042)	**0.011 (0.006)**	0.504 (0.059)	**0.037 (0.035)**	0.302 (0.042)	**0.011 (0.006)**
	U	0.299 (0.042)	**0.011 (0.006)**	0.504 (0.059)	**0.037 (0.036)**	0.302 (0.042)	**0.011 (0.006)**
	H	0.299 (0.042)	**0.011 (0.006)**	0.504 (0.059)	**0.037 (0.035)**	0.302 (0.042)	**0.011 (0.006)**
	G	0.299 (0.042)	**0.011 (0.006)**	0.504 (0.059)	**0.037 (0.036)**	0.302 (0.042)	**0.011 (0.006)**

Table S5: Confidence-based calibration metrics of `AutoGNNUQ-Simple` before and after recalibration based on the validation dataset. Mean value reported with standard deviation in parentheses.

Dataset	**Property**	**ECE**		**MCE**		**MCA**	
		Before	After	Before	After	Before	After
Lipo	logP	0.025 (0.011)	**0.012 (0.006)**	0.048 (0.017)	**0.026 (0.011)**	0.025 (0.011)	**0.012 (0.006)**
ESOL	S_{water}	0.040 (0.025)	**0.026 (0.013)**	0.082 (0.034)	**0.057 (0.019)**	0.041 (0.025)	**0.026 (0.013)**
FreeSolv	ΔG_{hyd}	0.084 (0.051)	**0.037 (0.015)**	0.159 (0.071)	**0.079 (0.030)**	0.085 (0.051)	**0.037 (0.015)**
QM7	$\Delta H^{\circ}_{\text{atom}}$	0.271 (0.013)	**0.031 (0.006)**	0.488 (0.022)	**0.129 (0.011)**	0.274 (0.013)	**0.032 (0.006)**
QM9	μ	0.032 (0.019)	**0.007 (0.003)**	0.058 (0.027)	**0.018 (0.010)**	0.032 (0.020)	**0.007 (0.004)**
	α	0.094 (0.015)	**0.006 (0.002)**	0.147 (0.022)	**0.018 (0.005)**	0.095 (0.015)	**0.006 (0.002)**
	ϵ_{HOMO}	0.071 (0.017)	**0.005 (0.002)**	0.114 (0.025)	**0.012 (0.003)**	0.072 (0.017)	**0.005 (0.002)**
	ϵ_{LUMO}	0.097 (0.017)	**0.005 (0.002)**	0.153 (0.027)	**0.013 (0.005)**	0.098 (0.018)	**0.005 (0.002)**
	$\Delta\epsilon$	0.091 (0.016)	**0.007 (0.002)**	0.144 (0.026)	**0.017 (0.006)**	0.092 (0.017)	**0.007 (0.002)**
	$\langle R^2 \rangle$	0.085 (0.018)	**0.007 (0.003)**	0.130 (0.027)	**0.016 (0.005)**	0.086 (0.018)	**0.007 (0.003)**
	ZPVE	0.125 (0.029)	**0.007 (0.002)**	0.198 (0.046)	**0.023 (0.006)**	0.126 (0.029)	**0.007 (0.002)**
	c_v	0.104 (0.023)	**0.009 (0.002)**	0.164 (0.036)	**0.027 (0.007)**	0.105 (0.023)	**0.009 (0.003)**
	U_0	0.419 (0.013)	**0.006 (0.002)**	0.761 (0.029)	**0.017 (0.004)**	0.424 (0.013)	**0.006 (0.002)**
	U	0.420 (0.014)	**0.008 (0.005)**	0.762 (0.028)	**0.022 (0.008)**	0.424 (0.014)	**0.009 (0.006)**
	H	0.424 (0.008)	**0.008 (0.002)**	0.768 (0.023)	**0.024 (0.007)**	0.428 (0.008)	**0.008 (0.002)**
	G	0.416 (0.013)	**0.010 (0.008)**	0.757 (0.023)	**0.025 (0.013)**	0.420 (0.013)	**0.011 (0.008)**

S6 Confidence Curve and Comparison with Evidential Uncertainty Quantification

Table S6: Model error (RMSE) at various confidence percentile cutoffs. Mean values are reported with the standard error of the mean in parentheses. The better result is bold and underlined. The random split ratio of `AutoGNNUQ` is 8:1:1 to be consistent with the benchmark evidential UQ [1].

Dataset	AutoGNNUQ	Benchmark	Dataset	AutoGNNUQ	Benchmark
	0.61 (0.01)	**0.55 (0.02)**		0.73 (0.04)	**0.66 (0.02)**
	0.49 (0.01)	0.50 (0.01)		0.55 (0.03)	**0.44 (0.01)**
Lipo	**0.45 (0.02)**	0.51 (0.02)	ESOL	0.47 (0.04)	**0.35 (0.02)**
	0.46 (0.04)	0.53 (0.03)		0.51 (0.06)	**0.28 (0.02)**
	0.47 (0.06)	0.50 (0.04)		0.53 (0.07)	**0.22 (0.02)**
	1.45 (0.12)	**0.96 (0.07)**		**107 (3)**	115 (3))
	0.86 (0.06)	**0.42 (0.04)**		**25 (4)**	39 (3)
FreeSolv	0.85 (0.08)	**0.26 (0.04)**	QM7	**7 (1)**	23 (4)
	0.85 (0.10)	**0.25 (0.08)**		**5 (1)**	10 (4)
	0.79 (0.17)	**0.28 (0.12)**		**5 (1)**	10 (4)

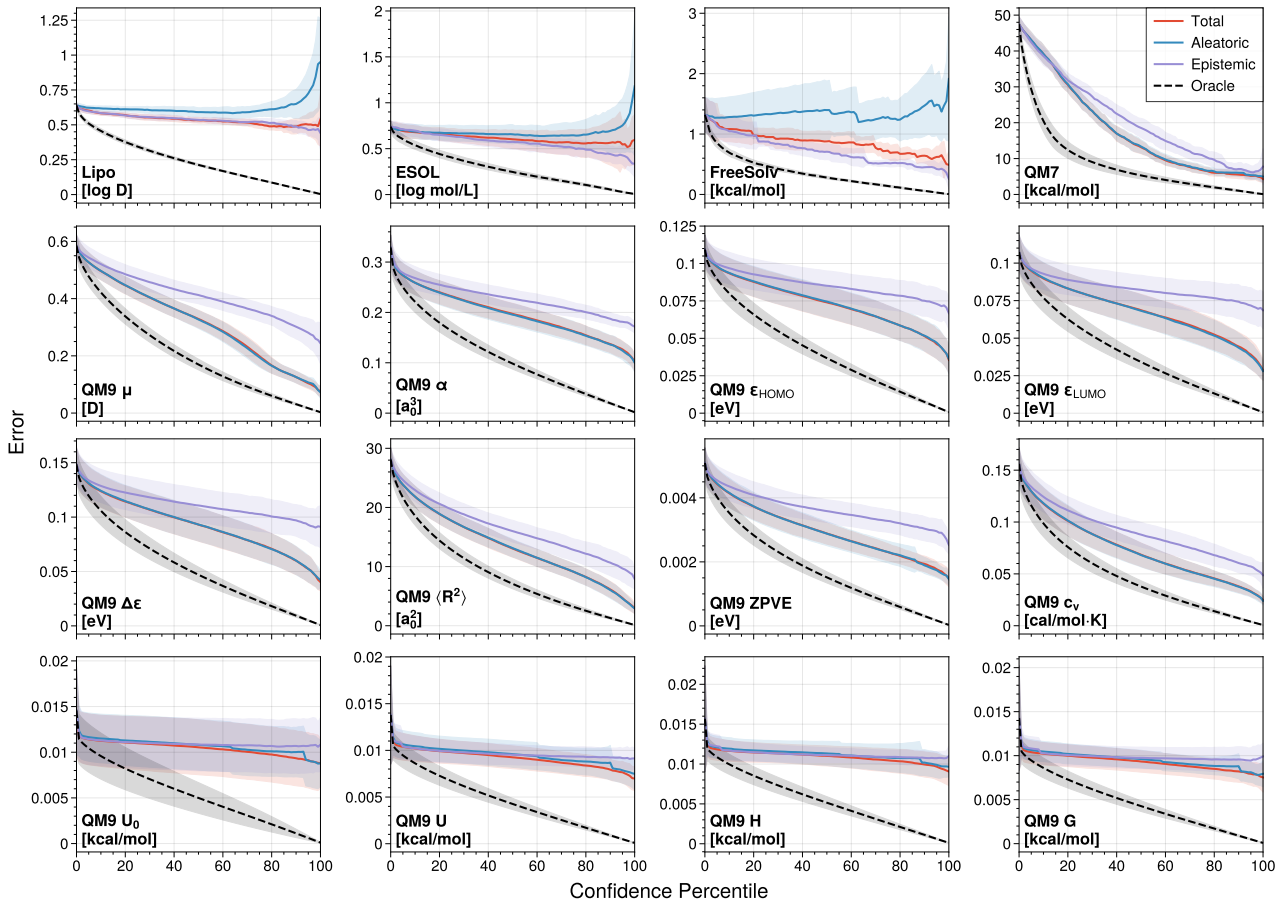

Figure S32: Confidence curve of `AutoUQGNN` of all test datasets across eight random seeds.

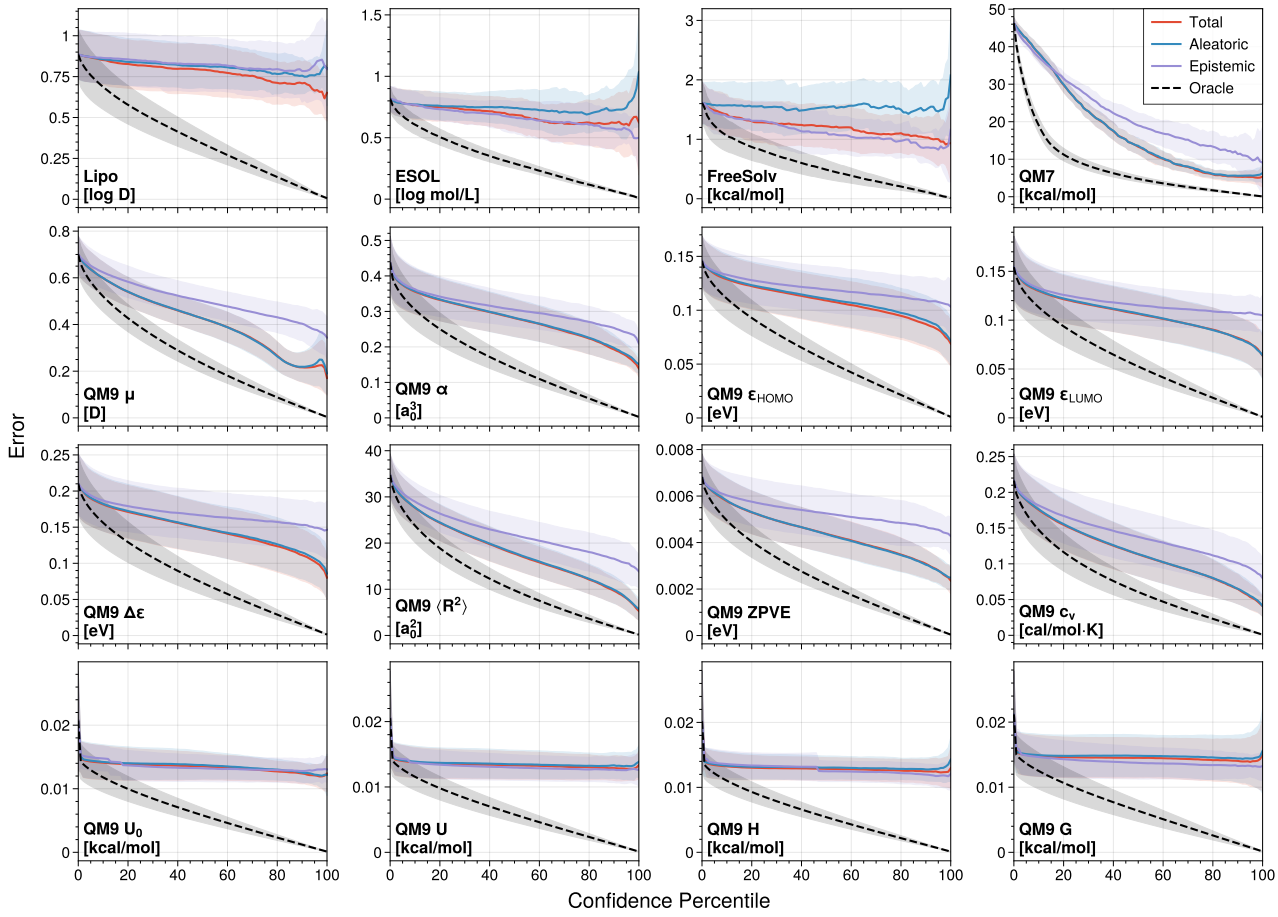

Figure S33: Confidence curve of `AutoUQGNN-Simple` of all test datasets across eight random seeds.

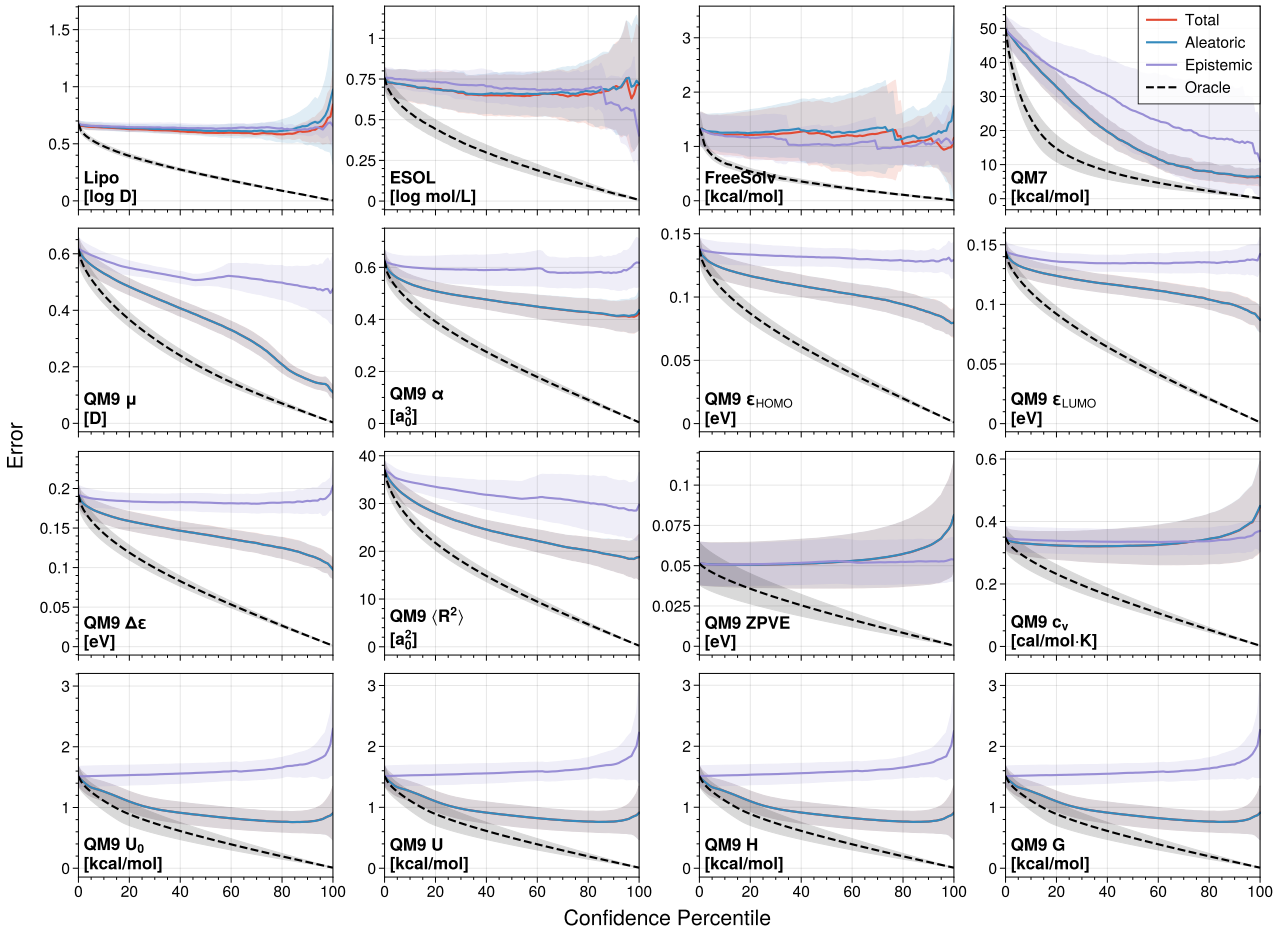

Figure S34: Confidence curve of MC dropout of all test datasets across eight random seeds.

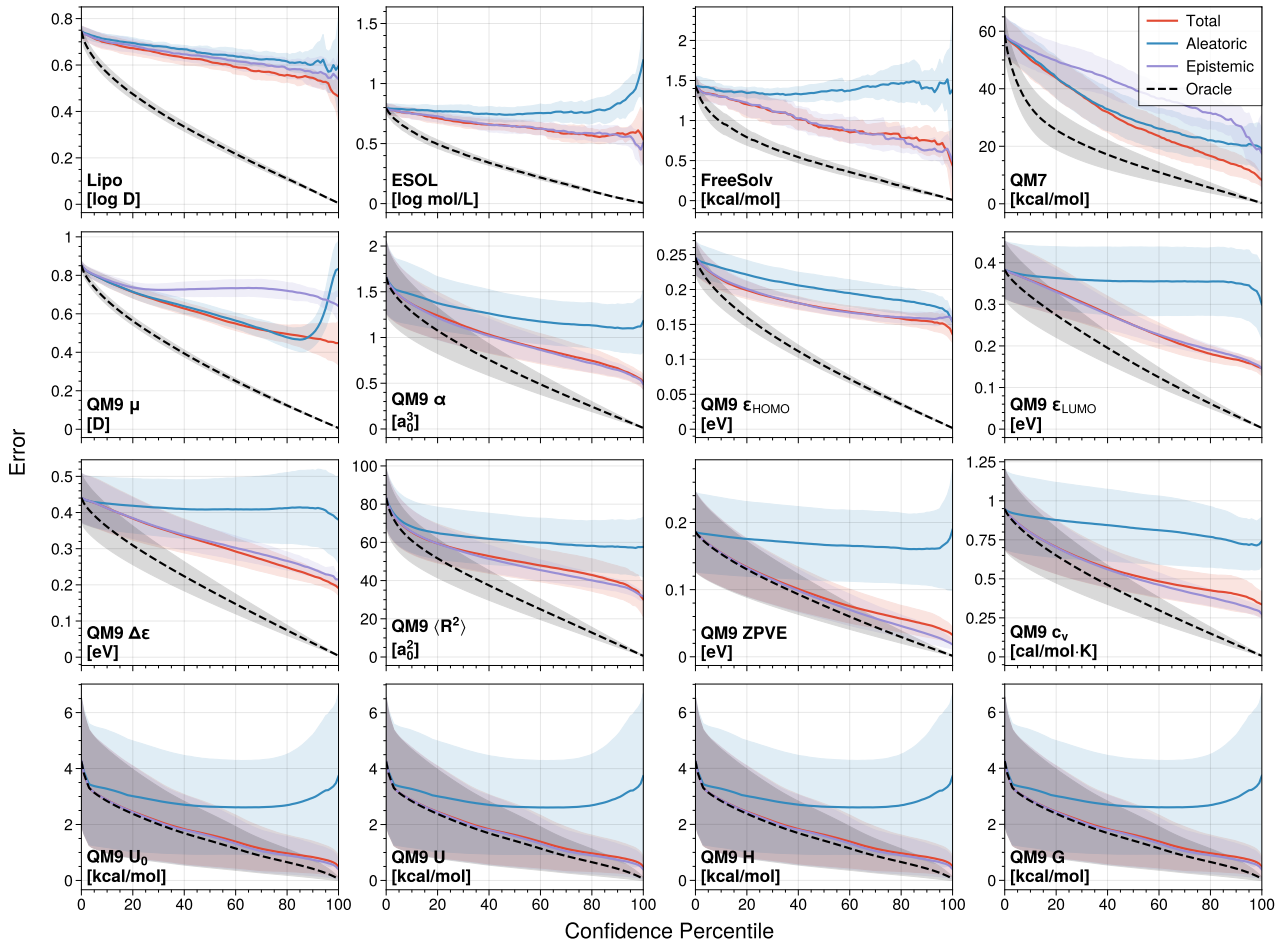

Figure S35: Confidence curve of random ensemble of all test datasets across eight random seeds.

Table S7: Confidence curve area under confidence-oracle error (AUCO). Mean value reported with standard deviation in parentheses. The best result is bold and underlined.

Dataset	Property	AutoGNNUQ	AutoGNNUQ-Simple	MC Dropout	Random Ensemble
Lipo	logP	**30.55 (1.75)**	42.06 (6.92)	37.64 (3.07)	32.16 (1.66)
ESOL	S_{water}	**34.00 (8.98)**	37.46 (10.96)	40.54 (11.45)	34.67 (2.71)
FreeSolv	ΔG_{hyd}	53.13 (10.23)	66.41 (18.86)	86.05 (47.86)	**48.63 (9.66)**
QM7	$\Delta H^{\circ}_{\text{atom}}$	**939.45 (63.19)**	989.97 (86.88)	1016.44 (104.93)	1332.97 (191.13)
QM9	μ	**11.77 (1.10)**	15.87 (2.70)	13.82 (1.11)	26.11 (2.35)
	α	**8.82 (0.95)**	12.48 (2.22)	22.91 (2.56)	32.47 (4.82)
	ϵ_{HOMO}	**3.29 (0.52)**	5.13 (1.14)	5.20 (0.48)	8.14 (0.47)
	ϵ_{LUMO}	**2.97 (0.50)**	4.88 (1.09)	5.78 (0.47)	8.91 (0.86)
	$\Delta\epsilon$	**4.00 (0.72)**	6.81 (1.84)	6.89 (0.67)	12.38 (1.21)
	$\langle R^2 \rangle$	**496.94 (64.98)**	684.88 (145.02)	1092.04 (139.38)	1887.69 (293.54)
	ZPVE	**0.12 (0.01)**	0.19 (0.04)	3.27 (1.17)	1.45 (0.64)
	c_v	**2.76 (0.45)**	4.62 (1.33)	19.45 (4.55)	16.17 (2.95)
	U_0	**0.53 (0.14)**	0.72 (0.10)	39.02 (7.78)	20.86 (17.48)
	U	**0.47 (0.06)**	0.73 (0.10)	39.04 (7.81)	20.81 (17.27)
	H	**0.56 (0.07)**	0.72 (0.10)	39.02 (7.81)	20.80 (17.29)
	G	**0.48 (0.06)**	0.79 (0.16)	39.03 (7.80)	20.72 (17.10)

S7 Loss Curves

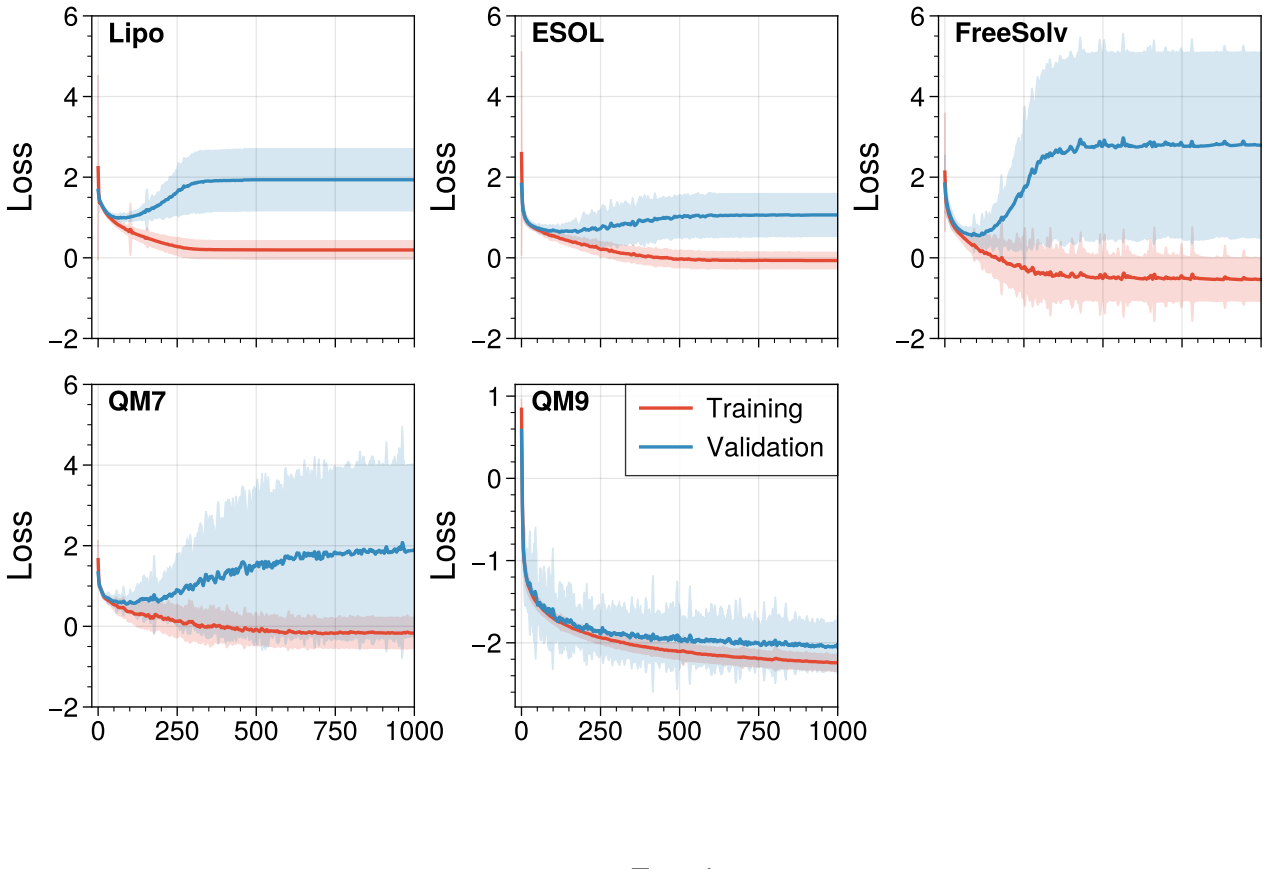

Figure S36: Loss curves of `AutoGNNUQ`. The solid line within the distribution represents the mean, while the shaded area signifies the standard deviation.

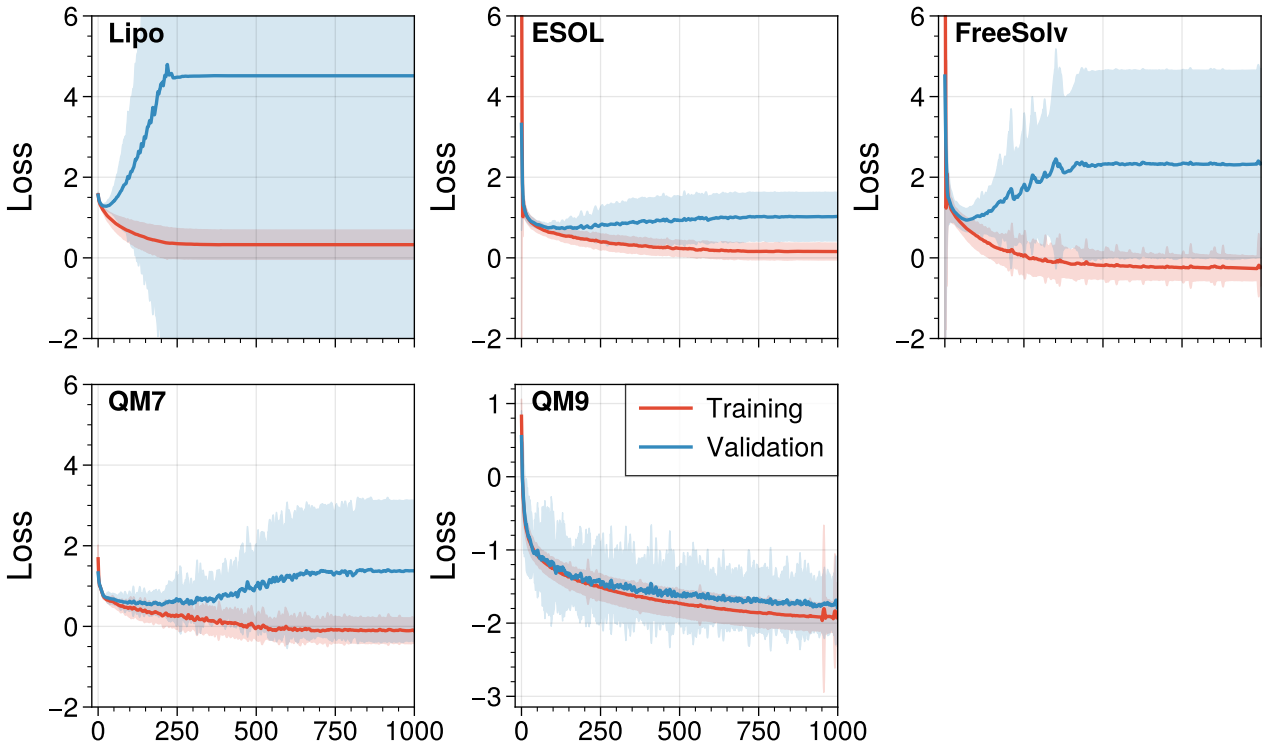

Figure S37: Loss curves of `AutoGNNUQ-Simple`. The solid line within the distribution represents the mean, while the shaded area signifies the standard deviation.

S8 Out-of-Distribution Performance

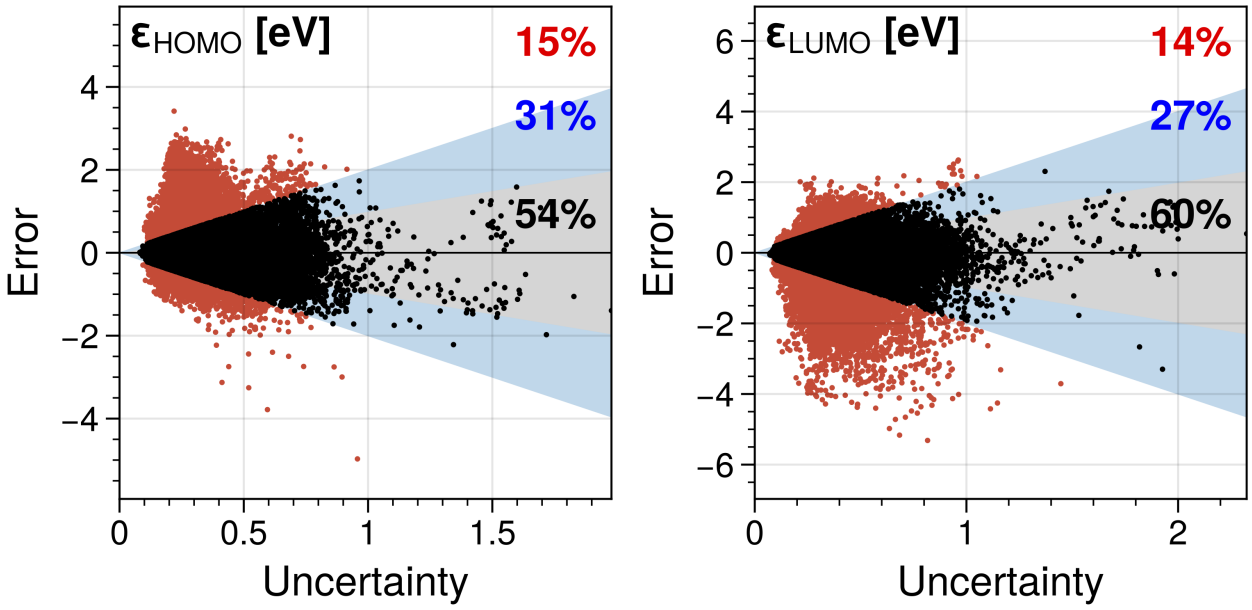

Figure S38: Prediction error vs uncertainty of PC9 using `AutoGNNUQ` trained on QM9.

References

[1] Soleimany, A. P. *et al.* Evidential deep learning for guided molecular property prediction and discovery. *ACS central science* **7**, 1356–1367 (2021).