Supporting Information: Revisiting a Large and Diverse Data Set for Barrier Heights and Reaction Energies: Best Practices in Density Functional Theory Calculations for Chemical Kinetics

Xiao Liu,[1] Kevin A. Spiekermann,[2] Angiras Menon,[2]
William H. Green[2] and Martin Head-Gordon*[,1,3]

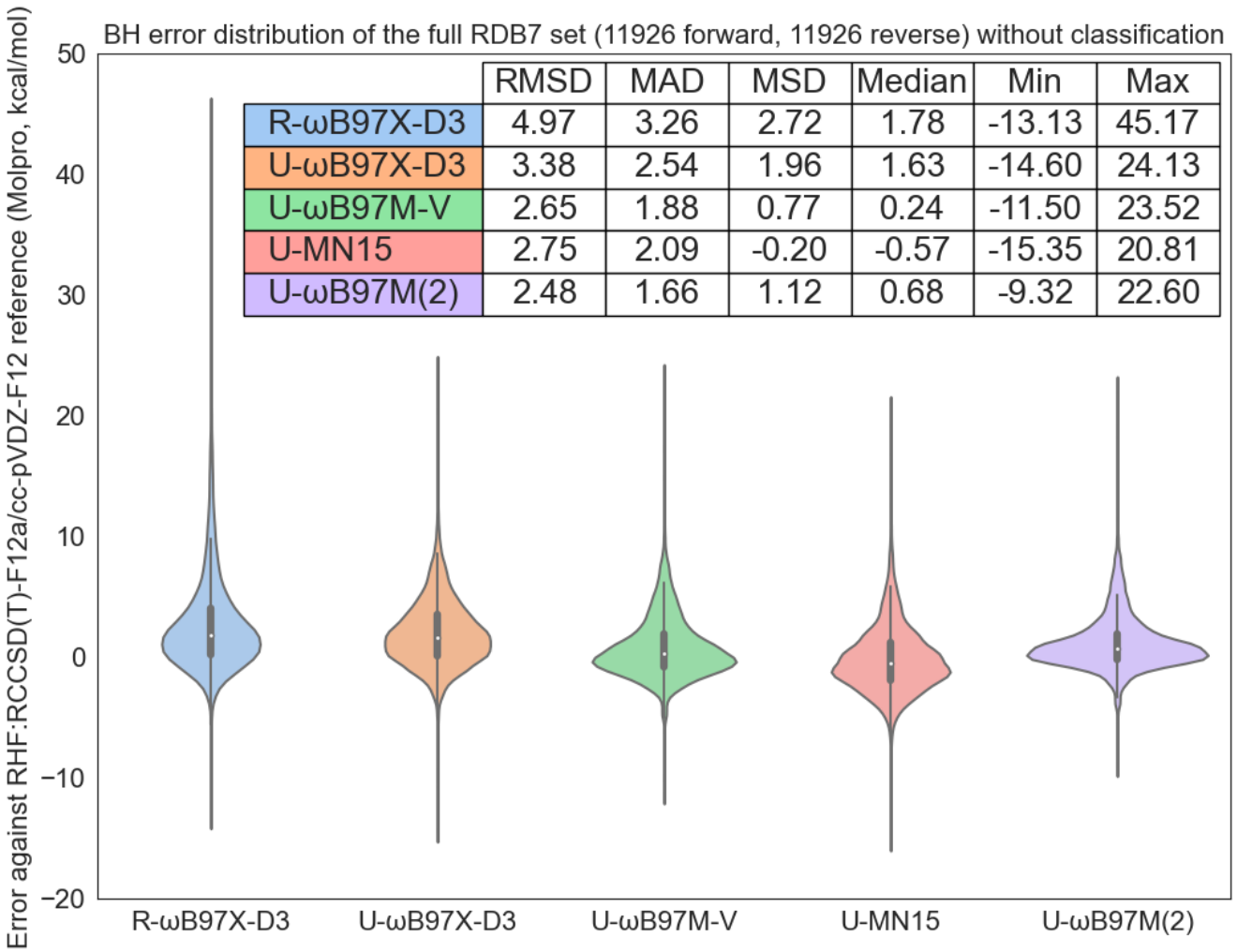
[1] *Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, University of California, Berkeley, California 94720, United States of America*

[2] *Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*

[3] *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States of America*

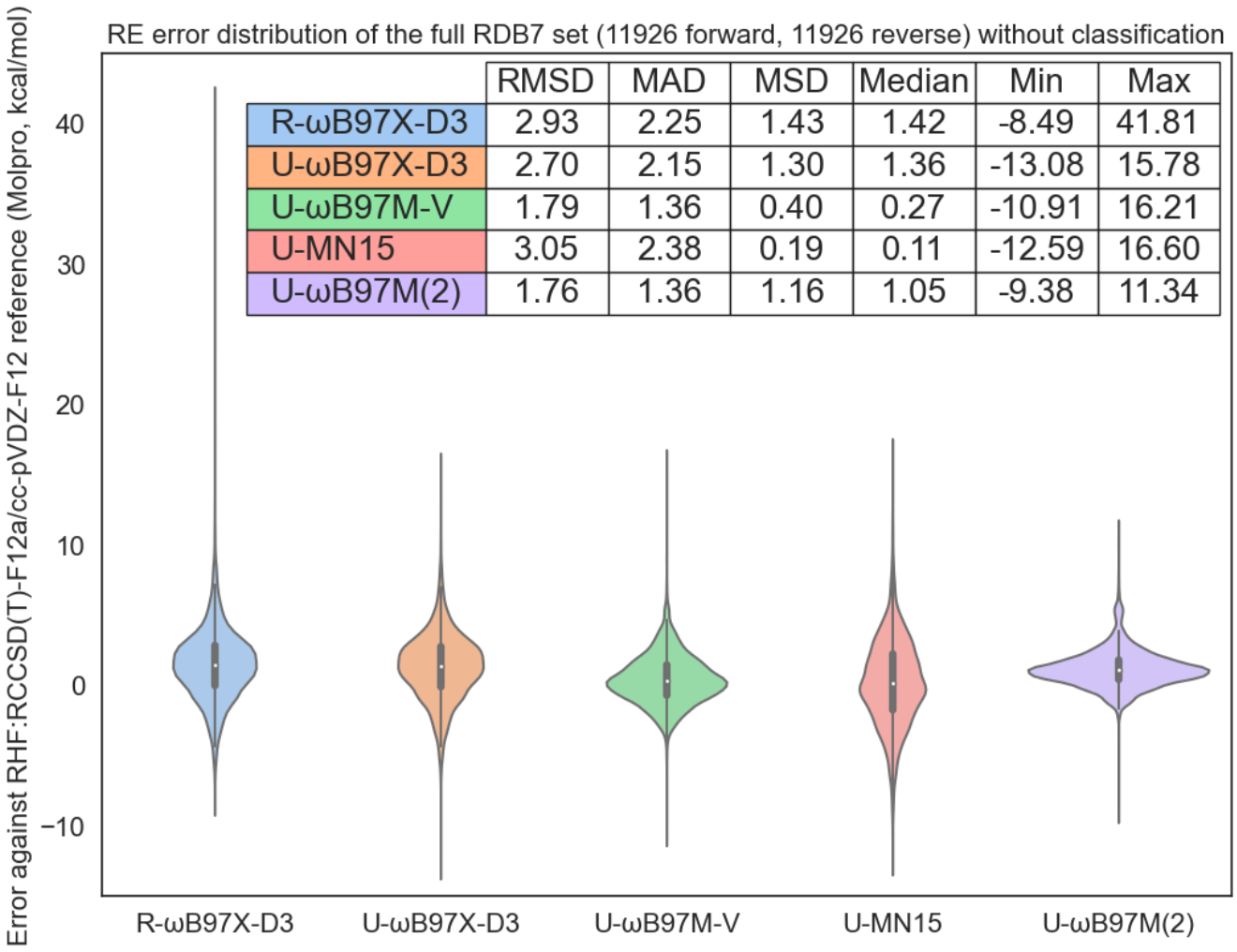
In the main text, we focus on the RMSD values of BH and RE data within each subset. Here, in the supporting information, we provide more detailed statistical analyses for each task. For reactions in the easy and intermediate subsets, which comprise the majority of the full RDB7 set, error statistics are presented in tables embedded within the violin plots. Comparisons between forward and reverse barrier heights are provided - explicitly for the easy and intermediate subsets using split violin plots, and implicitly for different collections of the refined reactions in the difficult subset. We also present the FBH, RBH, and RE error statistics for the 3rd collection of the difficult subset (50 data points with largest U-ωB97X-D3 deviations vs. existing RCCSD(T)-F12 reference). Finally, we include Q-Chem and ORCA input templates for the data categorization workflow of this work. For raw data of the full RDB7 set or within a specific subset, please refer to the corresponding subsheet in the RDB7_refined.xlsx file.

1 BH and RE error statistics of the uncategorized full RDB7 set



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	4.97	3.26	2.72	1.78	-13.13	45.17
U-ωB97X-D3	3.38	2.54	1.96	1.63	-14.60	24.13
U-ωB97M-V	2.65	1.88	0.77	0.24	-11.50	23.52
U-MN15	2.75	2.09	-0.20	-0.57	-15.35	20.81
U-ωB97M(2)	2.48	1.66	1.12	0.68	-9.32	22.60

Figure 1: Uncategorized full RDB7 set BH error distribution



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	2.93	2.25	1.43	1.42	-8.49	41.81
U-ωB97X-D3	2.70	2.15	1.30	1.36	-13.08	15.78
U-ωB97M-V	1.79	1.36	0.40	0.27	-10.91	16.21
U-MN15	3.05	2.38	0.19	0.11	-12.59	16.60
U-ωB97M(2)	1.76	1.36	1.16	1.05	-9.38	11.34

Figure 2: Uncategorized full RDB7 set RE error distribution

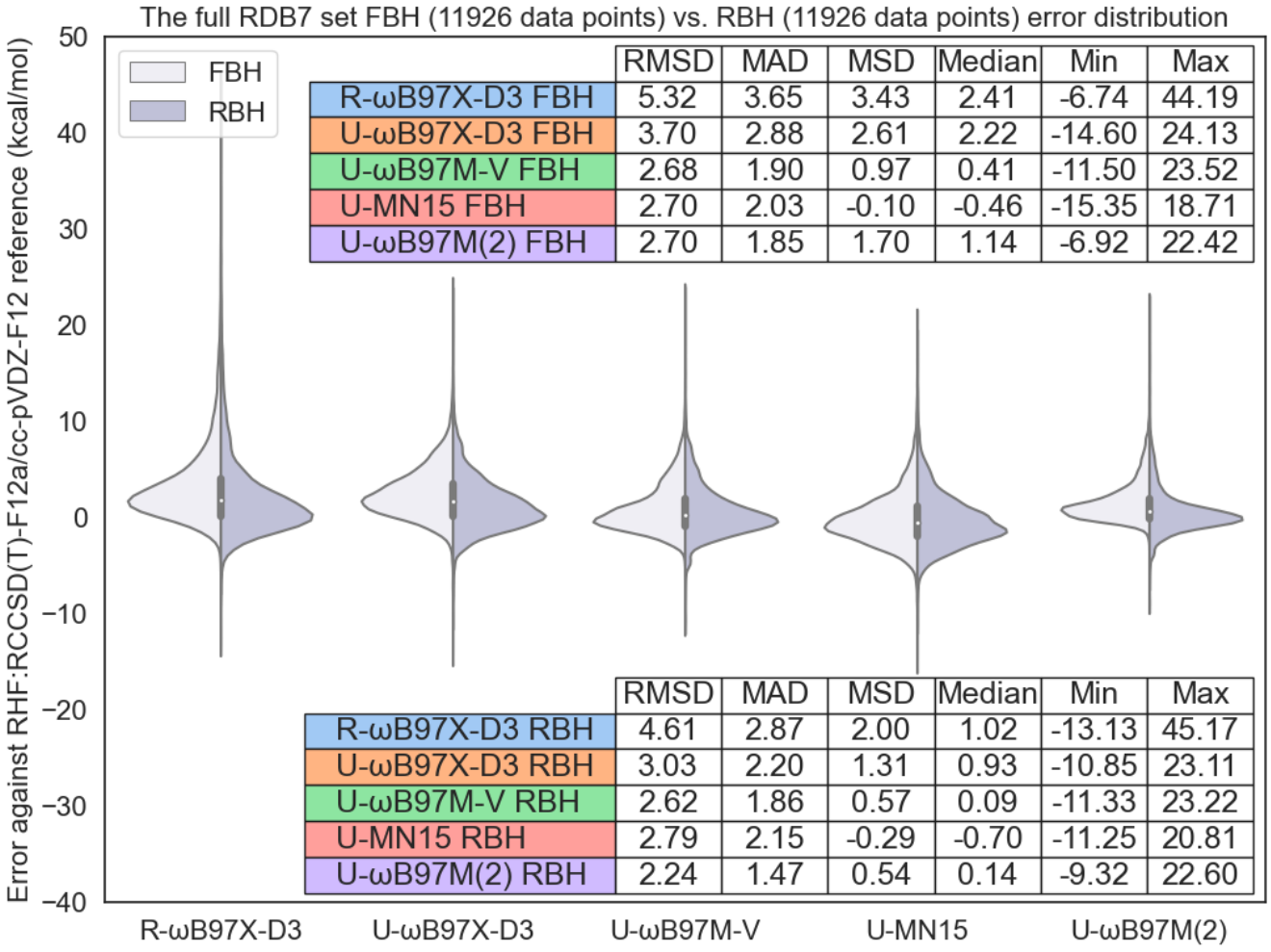
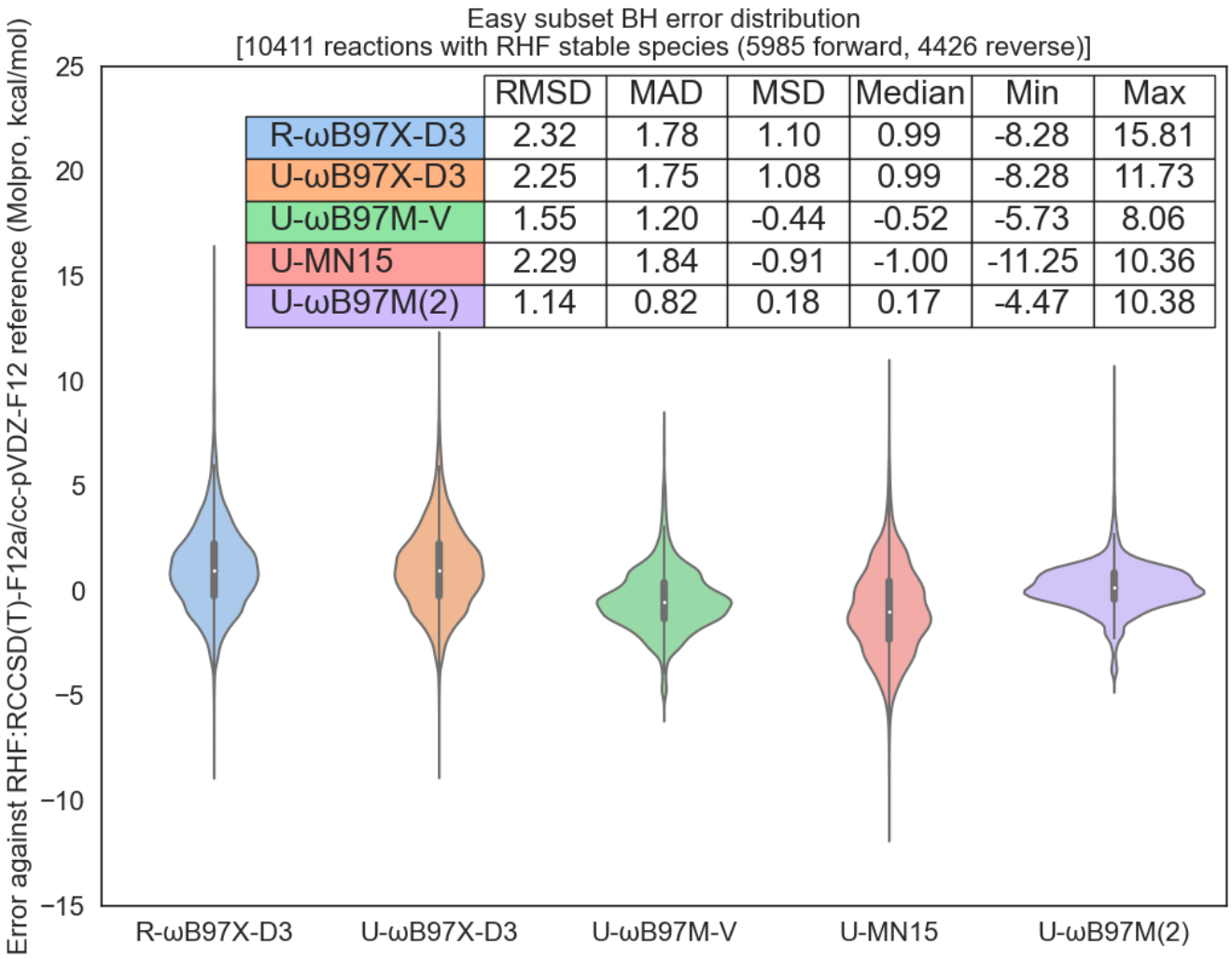


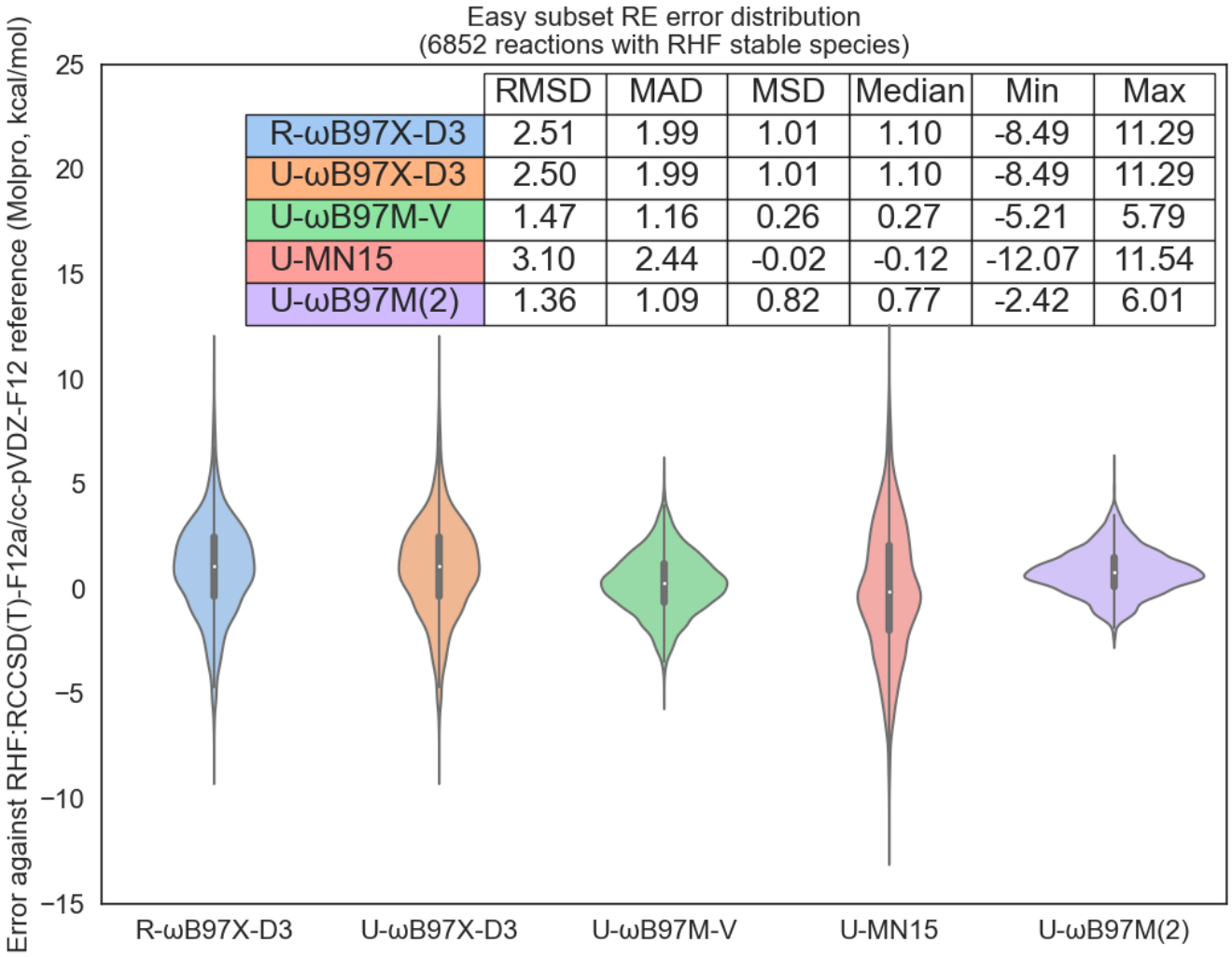
Figure 3: Uncategorized full RDB7 set FBH v.s. RBH error distribution

2 BH and RE error statistics of the easy subset



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	2.32	1.78	1.10	0.99	-8.28	15.81
U-ωB97X-D3	2.25	1.75	1.08	0.99	-8.28	11.73
U-ωB97M-V	1.55	1.20	-0.44	-0.52	-5.73	8.06
U-MN15	2.29	1.84	-0.91	-1.00	-11.25	10.36
U-ωB97M(2)	1.14	0.82	0.18	0.17	-4.47	10.38

Figure 4: Easy set BH error distribution



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	2.51	1.99	1.01	1.10	-8.49	11.29
U-ωB97X-D3	2.50	1.99	1.01	1.10	-8.49	11.29
U-ωB97M-V	1.47	1.16	0.26	0.27	-5.21	5.79
U-MN15	3.10	2.44	-0.02	-0.12	-12.07	11.54
U-ωB97M(2)	1.36	1.09	0.82	0.77	-2.42	6.01

Figure 5: Easy set RE error distribution

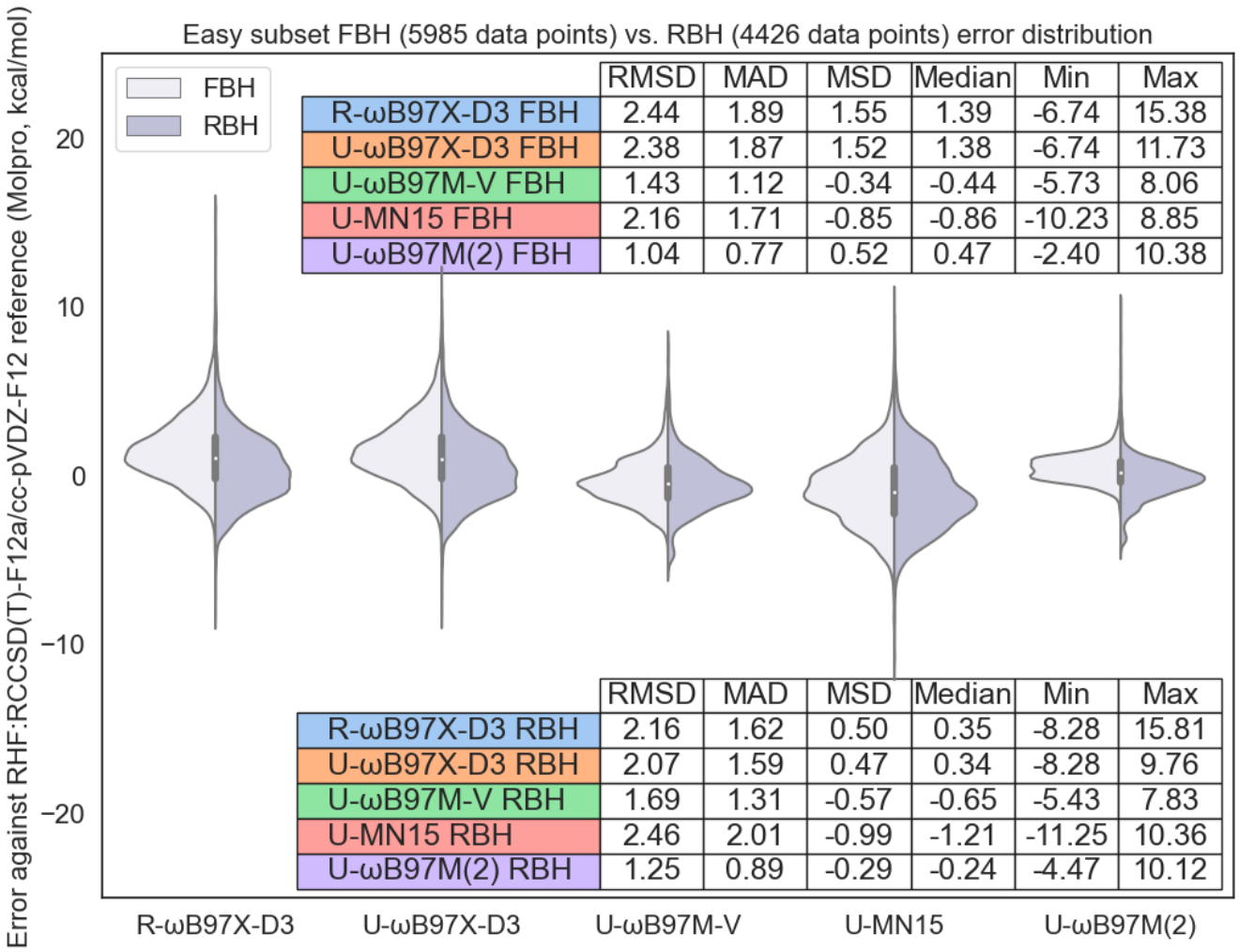
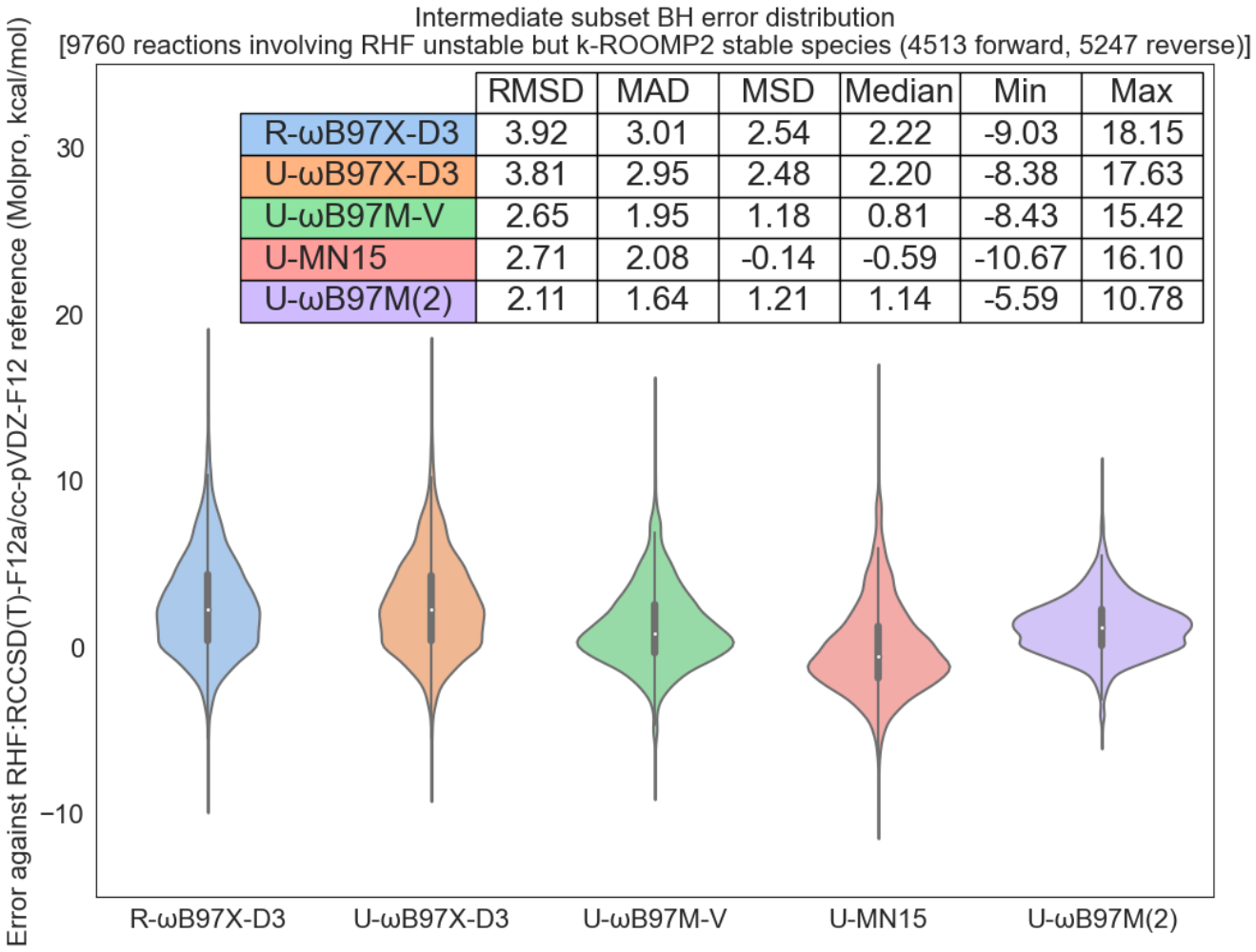


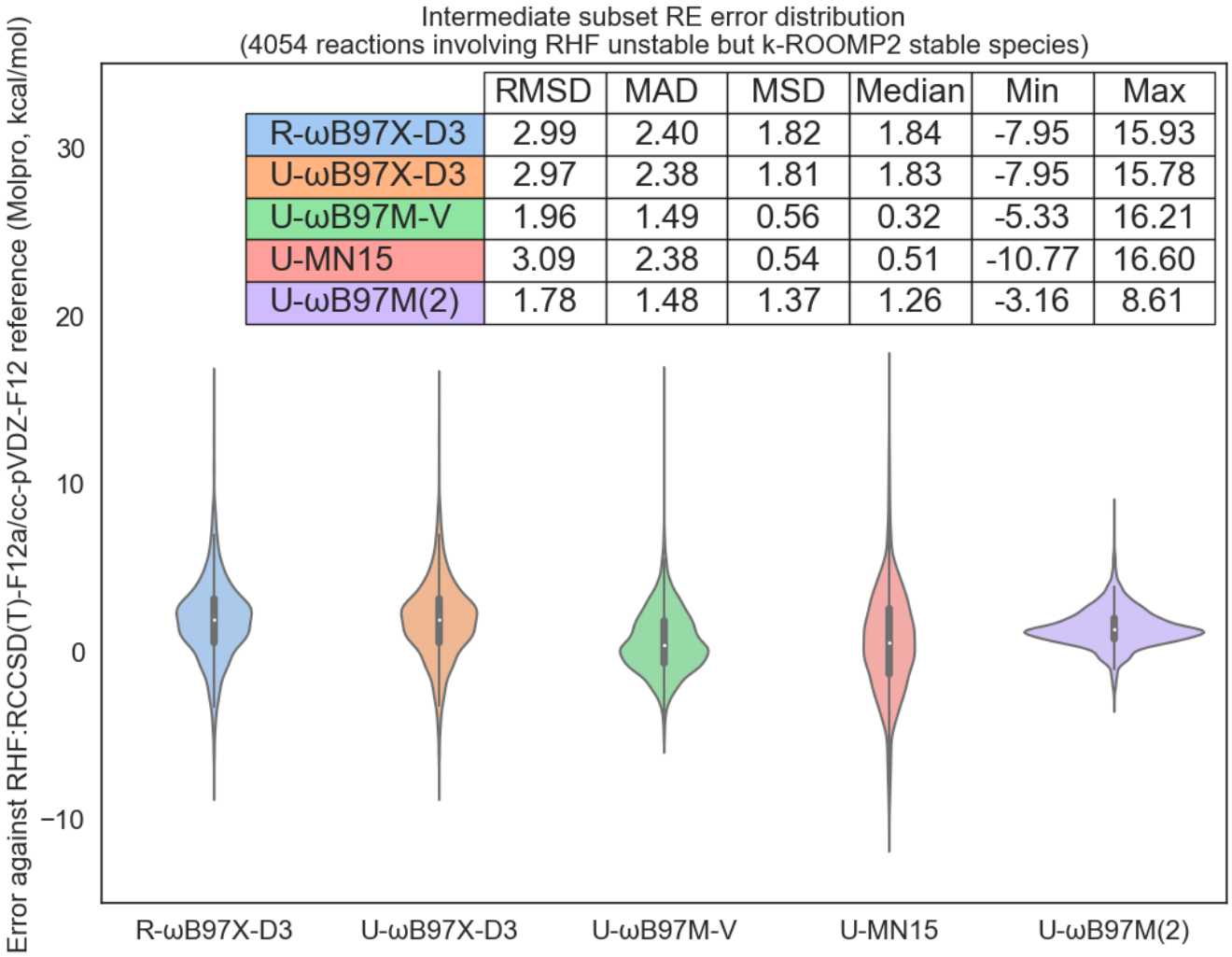
Figure 6: Easy set FBH v.s. RBH error distribution

3 BH and RE error statistics of the intermediate subset



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	3.92	3.01	2.54	2.22	-9.03	18.15
U-ωB97X-D3	3.81	2.95	2.48	2.20	-8.38	17.63
U-ωB97M-V	2.65	1.95	1.18	0.81	-8.43	15.42
U-MN15	2.71	2.08	-0.14	-0.59	-10.67	16.10
U-ωB97M(2)	2.11	1.64	1.21	1.14	-5.59	10.78

Figure 7: Intermediate set BH error distribution



	RMSD	MAD	MSD	Median	Min	Max
R-ωB97X-D3	2.99	2.40	1.82	1.84	-7.95	15.93
U-ωB97X-D3	2.97	2.38	1.81	1.83	-7.95	15.78
U-ωB97M-V	1.96	1.49	0.56	0.32	-5.33	16.21
U-MN15	3.09	2.38	0.54	0.51	-10.77	16.60
U-ωB97M(2)	1.78	1.48	1.37	1.26	-3.16	8.61

Figure 8: Intermediate set RE error distribution

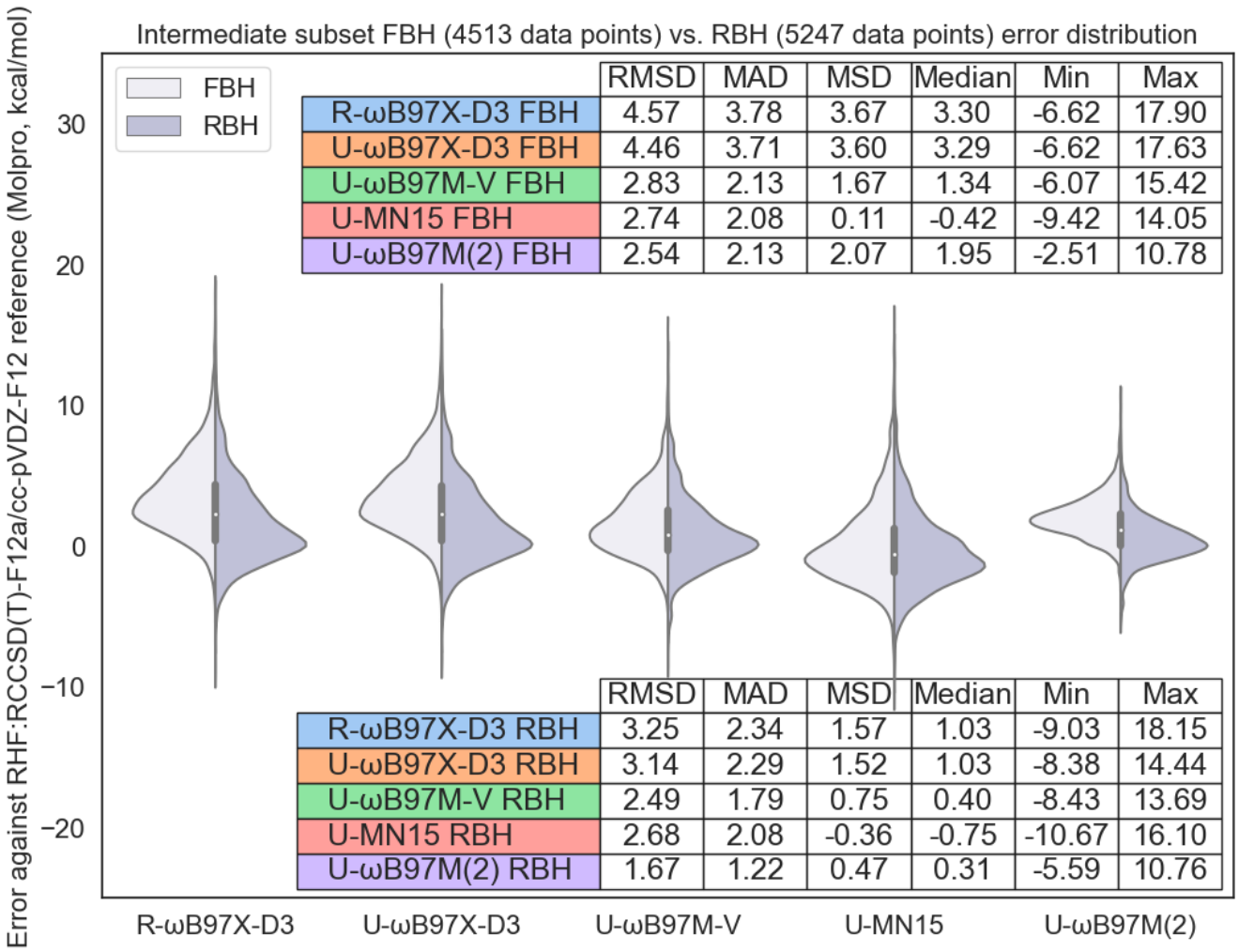


Figure 9: Intermediate set FBH v.s. RBH error distribution

4 1st collection in the difficult subset: RBH error statistics

Functional	Min	Max	MSD	MAD	RMSD
U-ωB97X-D3	-5.56	9.42	2.61	3.32	4.16
U-ωB97M-V	-3.78	8.88	4.12	4.27	4.86
U-MN15	-2.19	10.43	4.25	4.48	5.20
U-ωB97M(2)	-3.27	8.52	4.72	4.85	5.32

Table 1: RBH error statistics of unrestricted DFT compared against spin projected UHF:UCCSD(T)-F12 reference, for the 50 reactions with largest R-ωB97X-D3 v.s. RHF:RCCSD(T)-F12 RBH discrepancy (unit: kcal/mol)

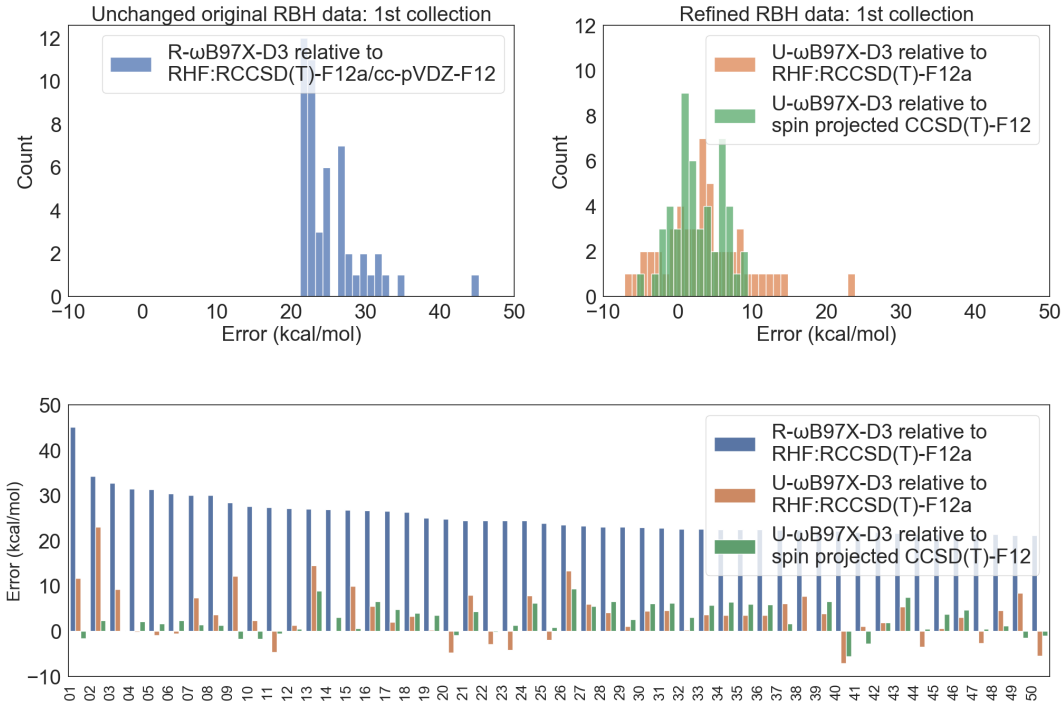


Figure 10: RBH error distribution for the 50 reactions with the largest R-ωB97X-D3 v.s. RHF:RCCSD(T)-F12 discrepancy.

5 2nd collection in the difficult subset: RBH error statistics

Functional	Min	Max	MSD	MAD	RMSD
U-ωB97X-D3	-3.34	14.40	10.29	10.56	11.03
U-ωB97M-V	-1.29	14.28	9.12	9.18	9.66
U-MN15	-2.23	15.68	8.12	8.26	9.14
U-ωB97M(2)	0.39	14.43	8.07	8.07	8.58

Table 2: RBH error statistics of unrestricted DFT compared against spin projected UHF:UCCSD(T)-F12 reference, for the 50 reactions with largest U-ωB97X-D3 v.s. RHF:RCCSD(T)-F12 RBH discrepancy (unit: kcal/mol)

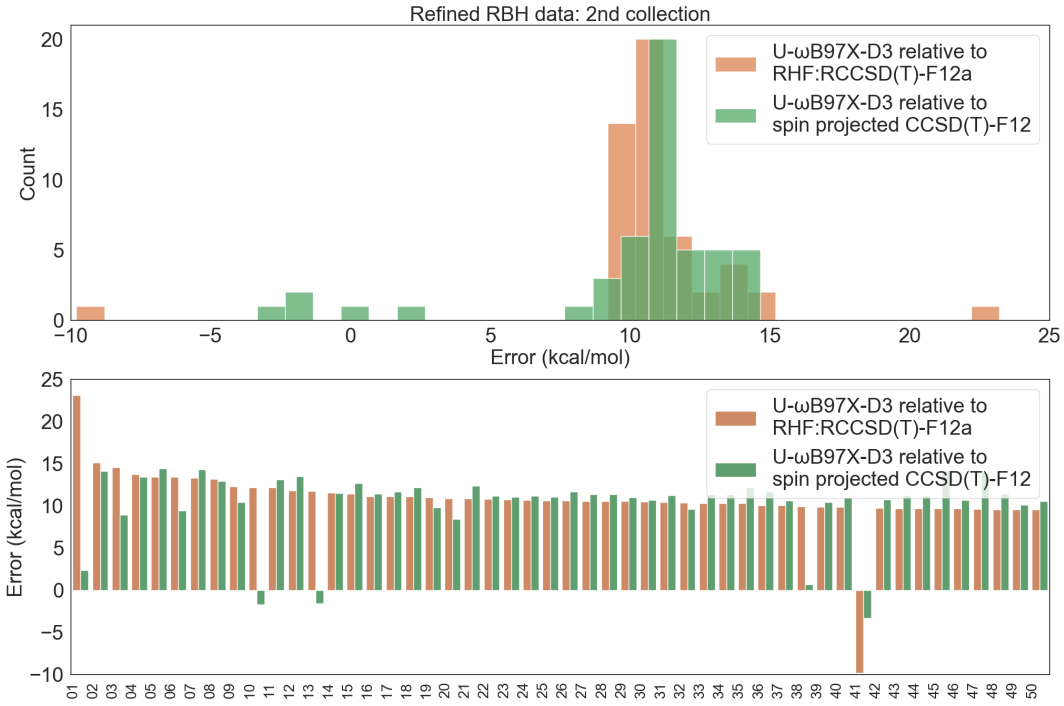


Figure 11: RBH error distribution for the 50 reactions with the largest U-ωB97X-D3 v.s. RHF:RCCSD(T)-F12 discrepancy.

6 3rd collection (largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 discrepancy) in the difficult subset: FBH, RBH, and RE error statistics

Functional	Min	Max	MSD	MAD	RMSD
U-ωB97X-D3	-3.24	17.26	8.73	9.02	10.00
U-ωB97M-V	0.20	14.61	8.41	8.41	9.19
U-MN15	-2.13	12.90	7.76	7.95	8.92
U-ωB97M(2)	1.58	12.74	8.23	8.23	8.74

Table 3: FBH error statistics of unrestricted DFT compared against spin projected UHF:UCCSD(T)-F12 reference, for the 50 reactions with largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 FBH discrepancy (unit: kcal/mol)

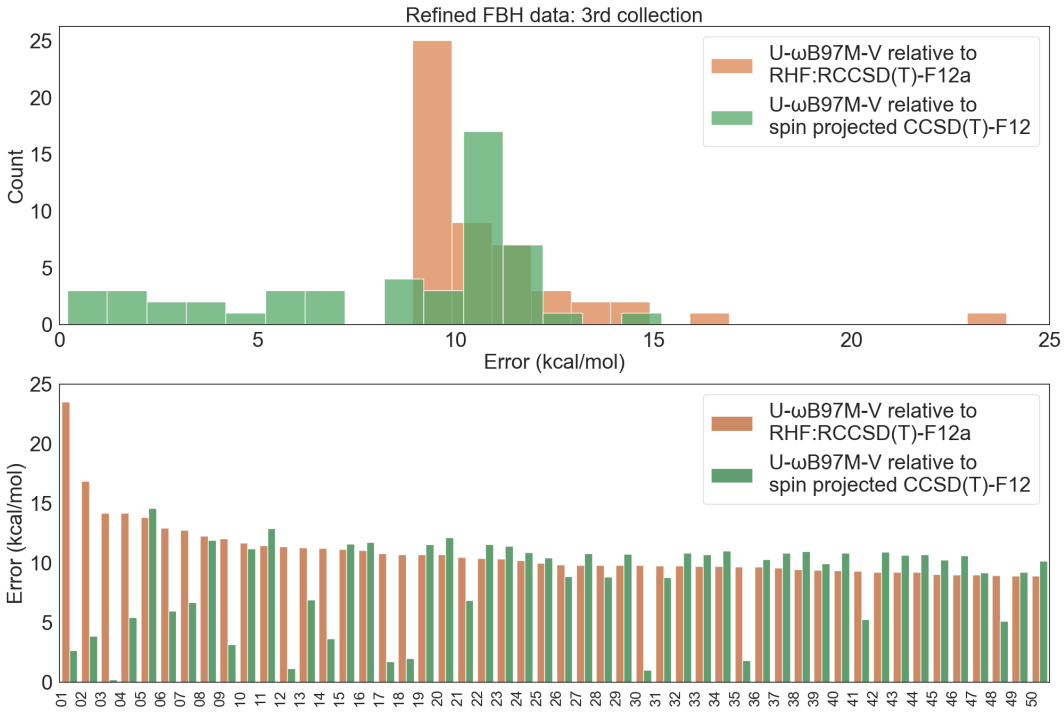


Figure 12: FBH error distribution for the 50 reactions with the largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 discrepancy.

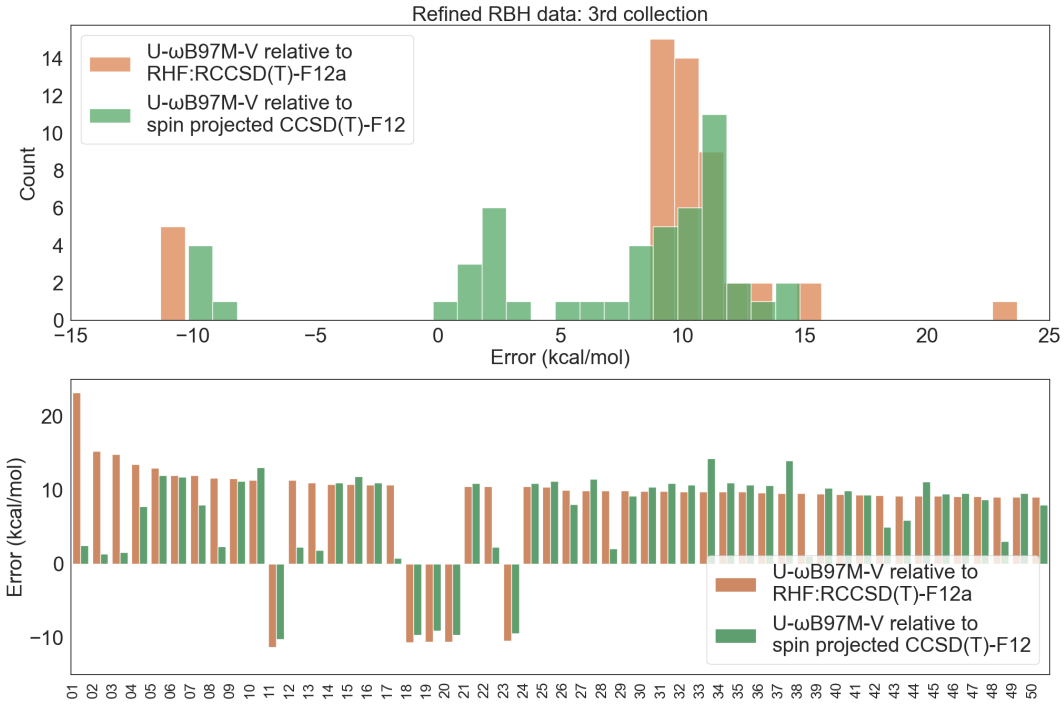


Figure 13: RBH error distribution for the 50 reactions with the largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 discrepancy.

Functional	Min	Max	MSD	MAD	RMSD
U-ωB97X-D3	-6.30	14.40	6.87	8.26	9.49
U-ωB97M-V	-10.21	14.28	6.44	8.36	9.18
U-MN15	-9.69	15.68	5.90	7.73	8.86
U-ωB97M(2)	-8.20	14.43	5.80	7.35	7.94

Table 4: RBH error statistics of unrestricted DFT compared against spin projected UHF:UCCSD(T)-F12 reference, for the 50 reactions with largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 RBH discrepancy (unit: kcal/mol)

Functional	Min	Max	MSD	MAD	RMSD
U-ωB97X-D3	1.68	13.06	7.07	7.07	7.35
U-ωB97M-V	4.25	11.30	7.40	7.40	7.49
U-MN15	0.75	9.46	5.51	5.51	5.88
U-ωB97M(2)	4.85	12.24	7.73	7.73	7.82

Table 5: RE error statistics of unrestricted DFT compared against spin projected UHF:UCCSD(T)-F12 reference, for the 50 reactions with largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 RE discrepancy (unit: kcal/mol)

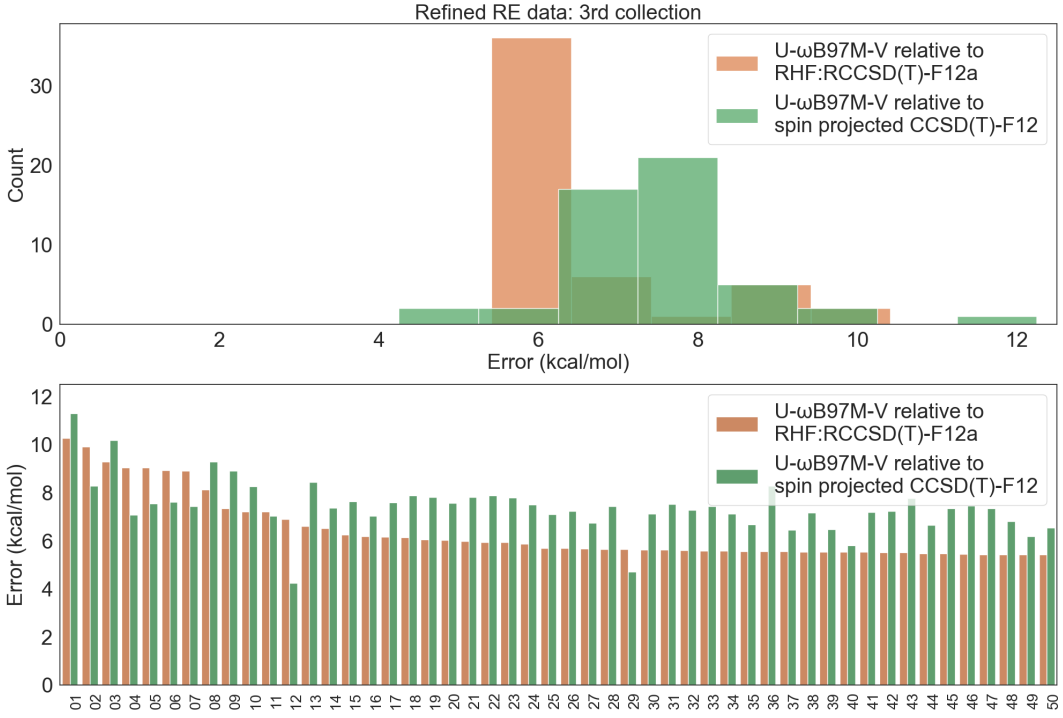


Figure 14: RE error distribution for the 50 reactions with the largest U-ωB97M-V v.s. RHF:RCCSD(T)-F12 discrepancy.

7 Sample input templates for data categorization workflow

7.1 Easy Subset: RHF stability analysis with Q-Chem

```
$rem
JOBTYPE                   SP
METHOD                    HF
BASIS                 cc-pVDZ
UNRESTRICTED            FALSE
SCF_ALGORITHM             GDM
MAX_SCF_CYCLES            100
SCF_CONVERGENCE             8
SYM_IGNORE               TRUE
SYMMETRY                FALSE
STABILITY_ANALYSIS       TRUE
$end
```

7.2 Intermediate Subset: unrestricted κ-OOMP2 $< S^2 >$ calculation with Q-Chem

```
$rem
JOBTYPE                   SP
METHOD                    HF
BASIS                 cc-pVDZ
UNRESTRICTED             TRUE
SCF_GUESS_MIX               1
SCF_ALGORITHM             GDM
MAX_SCF_CYCLES            100
SCF_CONVERGENCE             8
SYM_IGNORE               TRUE
SYMMETRY                FALSE
INTERNAL_STABILITY       TRUE
$end

@@@

$molecule
READ
$end

$rem
JOBTYPE                   SP
METHOD                 kOOMP2
REG_VARIABLE            1450 ! Set kappa = 1.45
BASIS                 cc-pVDZ
AUX_BASIS      RIMP2-cc-pVDZ
UNRESTRICTED             TRUE
SCF_GUESS                READ
SCF_ALGORITHM             GDM
MAX_SCF_CYCLES            100
SCF_CONVERGENCE             8
```

```
SYM_IGNORE              TRUE
SYMMETRY               FALSE
GEN_SCFMAN              TRUE
DO_S2                   TRUE ! Unrestricted kappa-OOMP2 <S^2> calculation
$end
```

7.3 Difficult Subset: UHF:UCCSD(T) and UCCSD $< S^2 >$ calculation with Q-Chem

```
$rem
JOBTYPE                   SP
METHOD                    HF
BASIS                cc-pVDZ
UNRESTRICTED             TRUE
SCF_GUESS_MIX               1
SCF_ALGORITHM             GDM
MAX_SCF_CYCLES            100
SCF_CONVERGENCE             8
SYM_IGNORE               TRUE
SYMMETRY                FALSE
INTERNAL_STABILITY       TRUE
$end

@@@

$molecule
READ
$end

$rem
JOBTYPE                   SP
METHOD                CCSD(T)
BASIS                 cc-pVDZ
UNRESTRICTED             TRUE
SCF_GUESS                READ
SCF_ALGORITHM             GDM
MAX_SCF_CYCLES            100
SCF_CONVERGENCE             8
SYM_IGNORE               TRUE
SYMMETRY                FALSE
GEN_SCFMAN               TRUE
CC_REF_PROP_TE           TRUE ! UCCSD <S^2> calculation
MEM_TOTAL                8000
$end
```

7.4 Selected Reactions from the Difficult Subset: UHF:UCCSD(T)-F12 calculation with ORCA

```
! CCSD(T)-F12/RI cc-pVDZ-F12 cc-pVDZ-F12-CABS cc-pVTZ/C # For restricted species
```

```
! UHF cc-pVDZ-F12 cc-pVDZ-F12-CABS # For unrestricted species
%scf Stabperform true end

$new_job
! CCSD(T)-F12/RI cc-pVDZ-F12 cc-pVDZ-F12-CABS cc-pVTZ/C
```