Supporting Information for

Machine learning to tame divergent density functional approximations: a new path to consensus materials design principles

Chenru Duan[1,2], Shuxin Chen[1], Michael G. Taylor[1], Fang Liu[1], and Heather J. Kulik[1]

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

[2]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Contents

Table S1. Summary of 23 functionals studied in this work, including their rungs on "Jacob's ladder" of DFT, Hartree–Fock (HF) exchange fraction, long-range correction (LRC) range-separation parameter ($bohr^{-1}$), MP2 correlation fraction, and whether empirical (i.e., D3) dispersion correction is included.

Functional	Type	Exchange type	HF exchange fraction	LRC range-separation parameter ($bohr^{-1}$)	MP2 correlation	D3 dispersion
BP86	GGA	GGA	--	--	--	No
BLYP	GGA	GGA	--	--	--	No
PBE	GGA	GGA	--	--	--	No
TPSS	meta-GGA	meta-GGA	--	--	--	No
SCAN	meta-GGA	meta-GGA	--	--	--	No
M06-L	meta-GGA	meta-GGA	--	--	--	No
MN15-L	meta-GGA	meta-GGA	--	--	--	No
B3LYP	GGA hybrid	GGA	0.200	--	--	No
B3P86	GGA hybrid	GGA	0.200	--	--	No
B3PW91	GGA hybrid	GGA	0.200	--	--	No
PBE0	GGA hybrid	GGA	0.250	--	--	No
ωB97X	RS hybrid	GGA	0.158	0.300	--	No
LRC-ωPBEh	RS hybrid	GGA	0.200	0.200	--	No
TPSSh	meta-GGA hybrid	meta-GGA	0.100	--	--	No
SCAN0	meta-GGA hybrid	meta-GGA	0.250	--	--	No
M06	meta-GGA hybrid	meta-GGA	0.270	--	--	No
M06-2X	meta-GGA hybrid	meta-GGA	0.540	--	--	No
MN15	meta-GGA hybrid	meta-GGA	0.440	--	--	No
B2GP-PLYP	double hybrid	GGA	0.650	--	0.360	No
PBE0-DH	double hybrid	GGA	0.500	--	0.125	No
DSD-BLYP-D3BJ	double hybrid	GGA	0.710	--	1.000	Yes
DSD-PBEB95-D3BJ	double hybrid	GGA	0.660	--	1.000	Yes
DSD-PBEP6-D3BJ	double hybrid	GGA	0.690	--	1.000	Yes

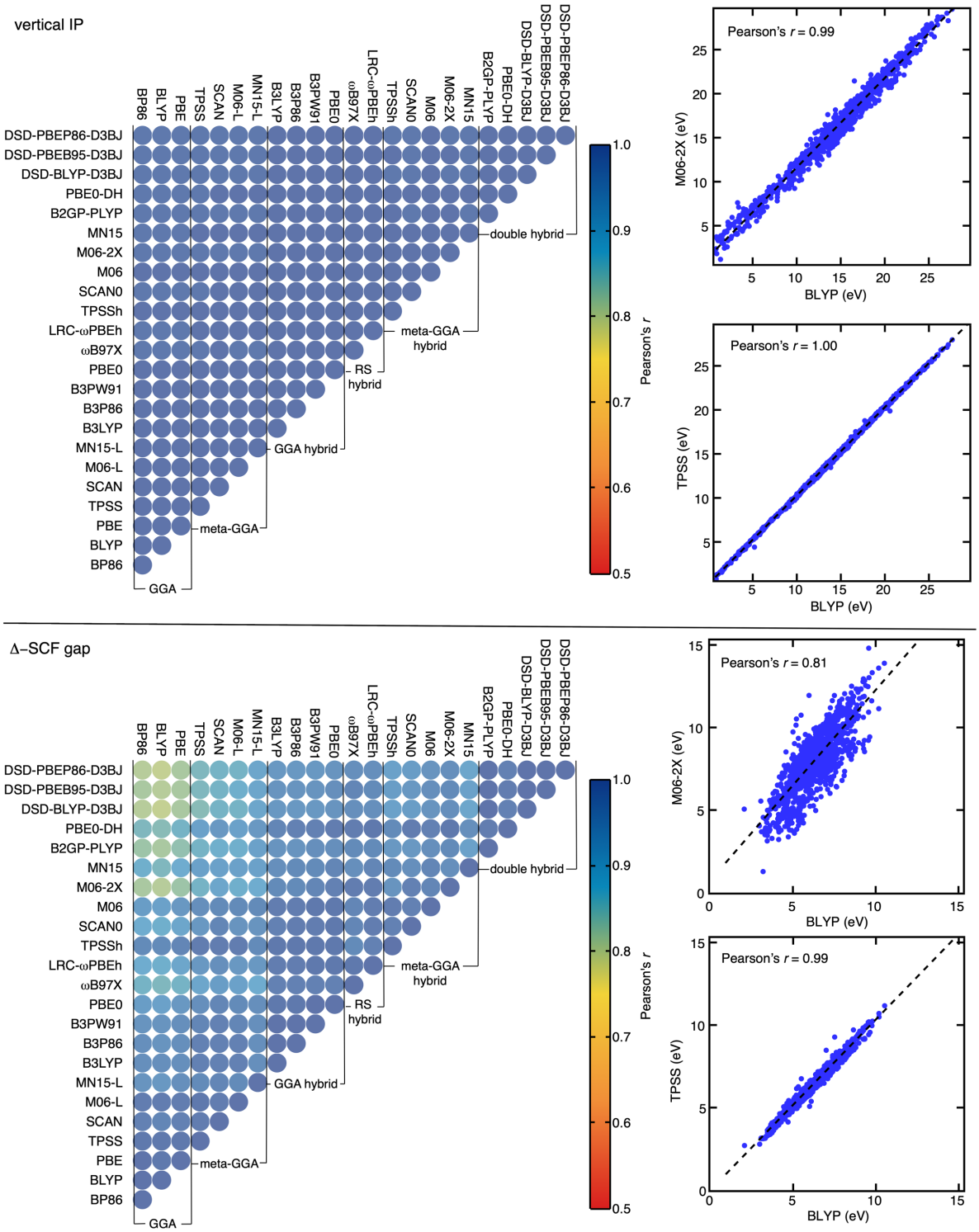

Figure S1. (top) An upper triangular matrix colored by Pearson's *r* for pairs of 23 functionals for vertical IP (shown at left) along with a parity plot of vertical IP for M06-2X and BLYP (top right) and for TPSS and BLYP (bottom right). (bottom) An upper triangular matrix colored by Pearson's *r* for pairs of 23 functionals for Δ-SCF gap (shown at left) along with a parity plot of Δ-SCF gap for M06-2X and BLYP (top right), and for TPSS and BLYP (bottom right).

Table S2. Summary of the mean value, standard deviation (std. dev.), minimum (min.), and maximum (max.) with each DFA for three properties considered in this work. Units of each measurement are indicated at top.

	$\Delta E_{H\text{-}L}$ (kcal/mol)				Δ-SCF gap (eV)				vertical IP (eV)			
	mean	std. dev.	min.	max.	mean	std. dev.	min.	max.	mean	std. dev.	min.	max.
BP86	11.4	30.8	-52.0	132.8	6.3	1.3	2.1	10.8	14.7	4.8	0.9	28.0
BLYP	7.3	26.5	-49.4	113.2	6.3	1.3	2.1	10.5	14.5	4.8	0.7	27.6
PBE	8.8	30.9	-56.5	133.3	6.3	1.3	2.2	10.8	14.6	4.8	0.7	27.8
TPSS	4.4	28.0	-52.8	116.0	6.5	1.4	2.7	11.2	14.8	4.8	0.7	28.0
SCAN	1.5	30.6	-66.0	110.4	6.8	1.5	2.0	11.8	15.0	4.8	0.7	28.6
M06-L	-2.2	26.4	-59.2	100.4	6.7	1.4	2.5	11.8	15.0	4.8	0.9	28.4
MN15-L	-27.5	33.8	-98.7	82.5	7.0	1.5	2.2	12.2	15.1	4.8	1.1	28.6
B3LYP	-8.6	24.5	-60.5	80.2	7.1	1.5	2.5	13.8	15.3	4.9	1.2	28.7
B3P86	-6.0	26.8	-61.5	90.4	7.1	1.5	2.2	14.0	15.6	4.8	1.2	29.0
B3PW91	-10.0	27.1	-66.5	87.1	7.2	1.5	2.1	14.0	15.4	4.8	1.0	28.8
PBE0	-14.6	27.0	-72.3	82.2	7.3	1.6	1.8	14.4	15.5	4.9	1.0	29.0
ωB97X	-12.3	24.4	-60.6	79.3	7.8	1.7	0.9	14.6	15.9	4.9	1.5	29.2
LRC-ωPBEh	-13.5	27.8	-69.1	79.3	7.6	1.7	1.8	14.5	15.7	4.9	1.1	29.1
TPSSh	-5.0	26.5	-60.3	94.1	6.5	1.9	2.6	11.9	15.1	4.8	0.7	28.4
SCAN0	-19.0	27.3	-82.0	74.7	7.6	1.8	0.9	15.3	15.7	4.9	0.9	29.6
M06	-21.2	29.8	-83.3	80.4	7.2	1.5	1.9	13.8	15.5	4.8	1.3	29.0
M06-2X	-40.9	26.3	-122.8	50.7	8.0	1.9	1.3	14.8	16.2	4.9	1.2	29.6
MN15	-7.9	29.6	-75.6	90.6	7.0	1.6	2.3	13.0	15.6	4.9	1.3	29.1
B2GP-PLYP	-20.5	28.3	-71.9	85.1	7.8	1.8	1.9	14.5	15.6	4.9	1.0	28.5
PBE0-DH	-22.5	28.0	-80.3	80.6	7.8	1.8	1.6	15.0	15.8	4.9	1.2	29.2
DSD-BLYP-D3BJ	-20.4	28.7	-69.9	89.1	7.9	1.9	1.9	14.5	15.6	5.0	0.9	28.5
DSD-PBEB95-D3BJ	-19.7	28.1	-69.2	87.1	7.9	1.8	2.0	14.4	15.6	4.9	1.1	28.6
DSD-PBEP6-D3BJ	-20.5	29.0	-70.7	90.6	7.9	1.8	1.9	14.5	15.6	5.0	0.8	28.5

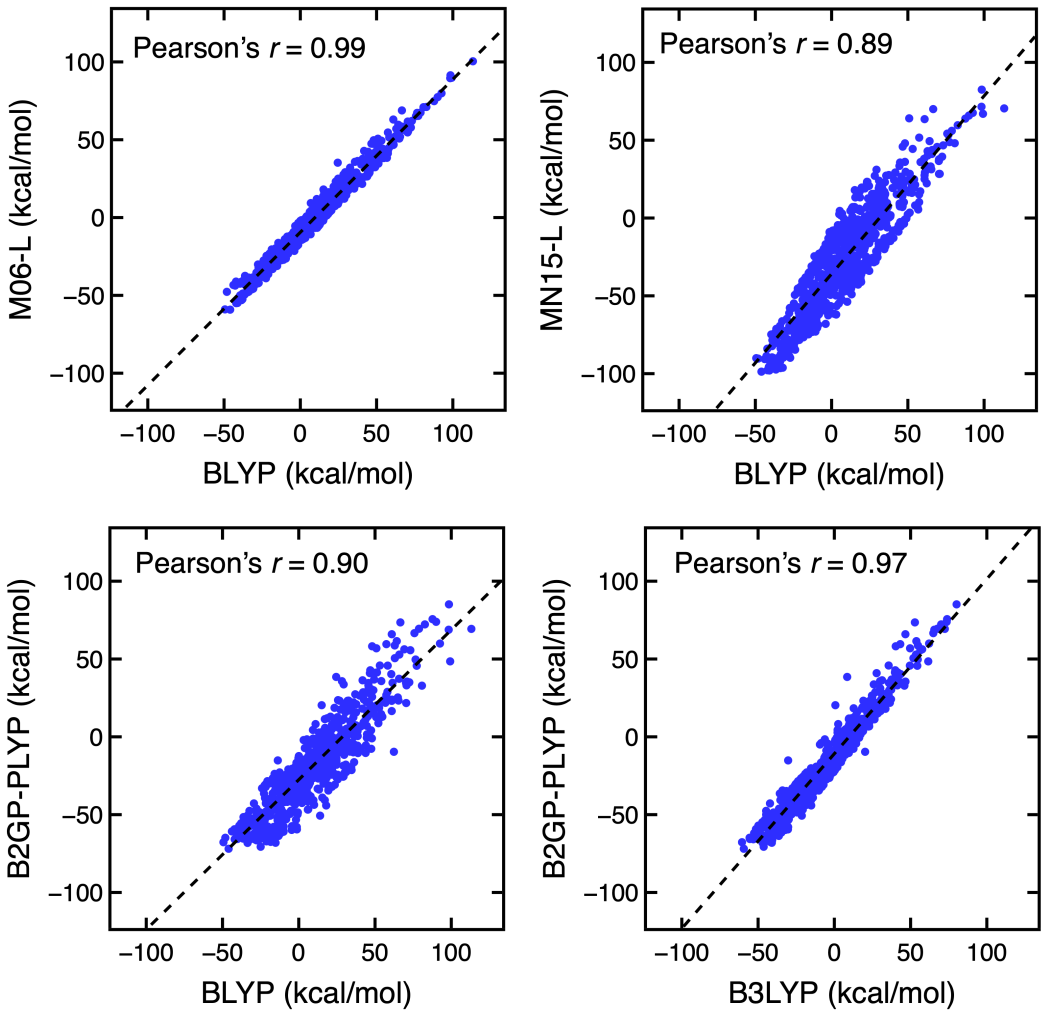

Figure S2. Parity plots of $\Delta E_{\text{H-L}}$ for select functionals: BLYP vs. M06-L (top left), BLYP vs. MN15-L (top right), BLYP vs B2GP-PLYP (bottom left), and B3LYP vs B2GP-PLYP (bottom right). A linear regression fit is shown as black dashed line in each parity plot, and the Pearson's *r* is shown in inset.

Table S3. Pearson's *r* of $\Delta E_{\text{H-L}}$ among DFAs from the M06 family using the LACVP* basis.

	M06-L	M06	M06-2X
M06-L	1.00	0.95	0.72
M06	0.95	1.00	0.86
M06-2X	0.72	0.86	1.00

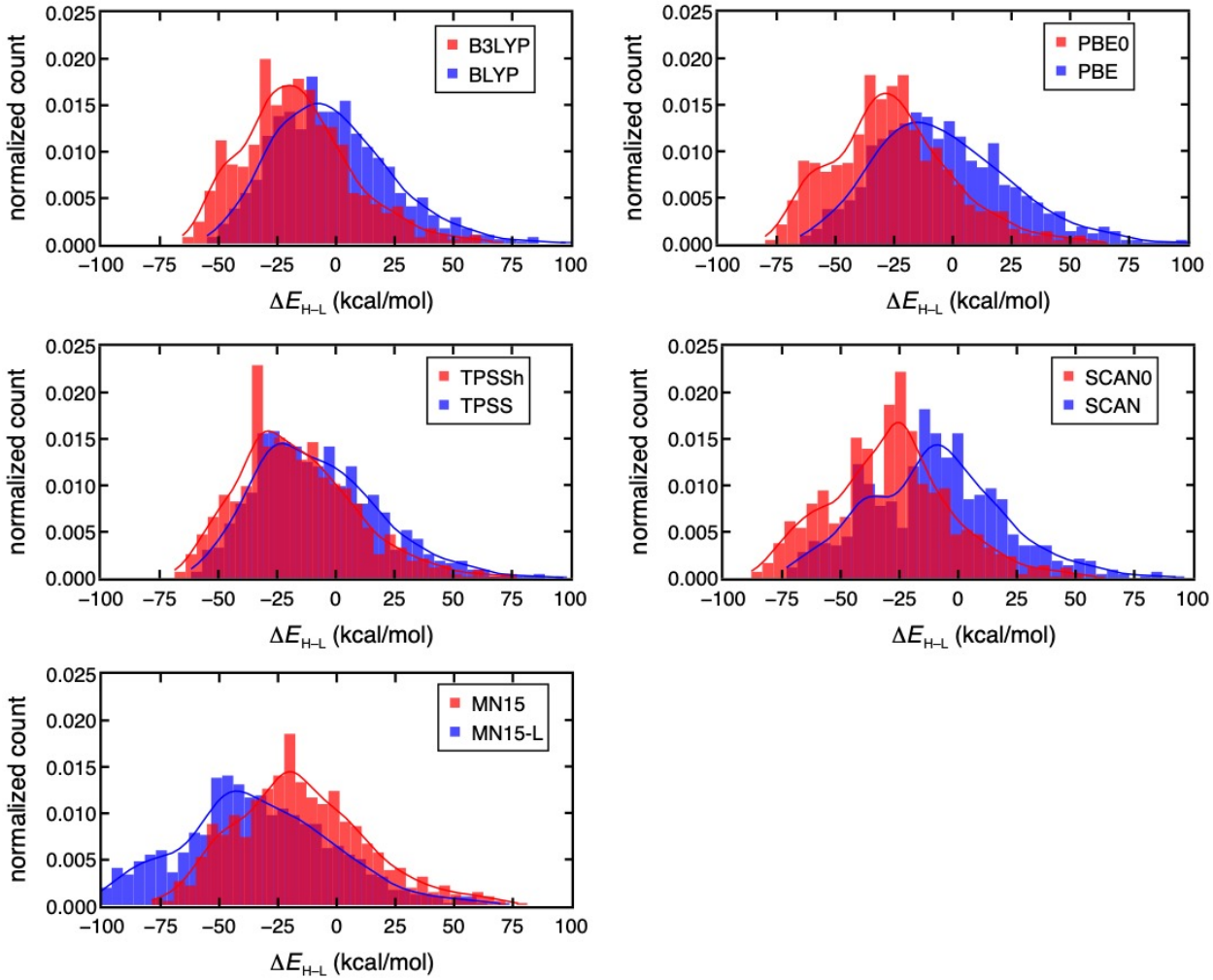

Figure S3. Distribution of $\Delta E_{\text{H-L}}$ (in kcal/mol) for select DFAs (from left to right and top to bottom, as indicated in inset legend): BLYP and B3LYP (top left), PBE and PBE0 (top right), TPSS and TPSSh (middle left), SCAN and SCAN0 (middle right), and MN15-L and MN15 (bottom left). The bin size is 5 kcal/mol for all distributions.

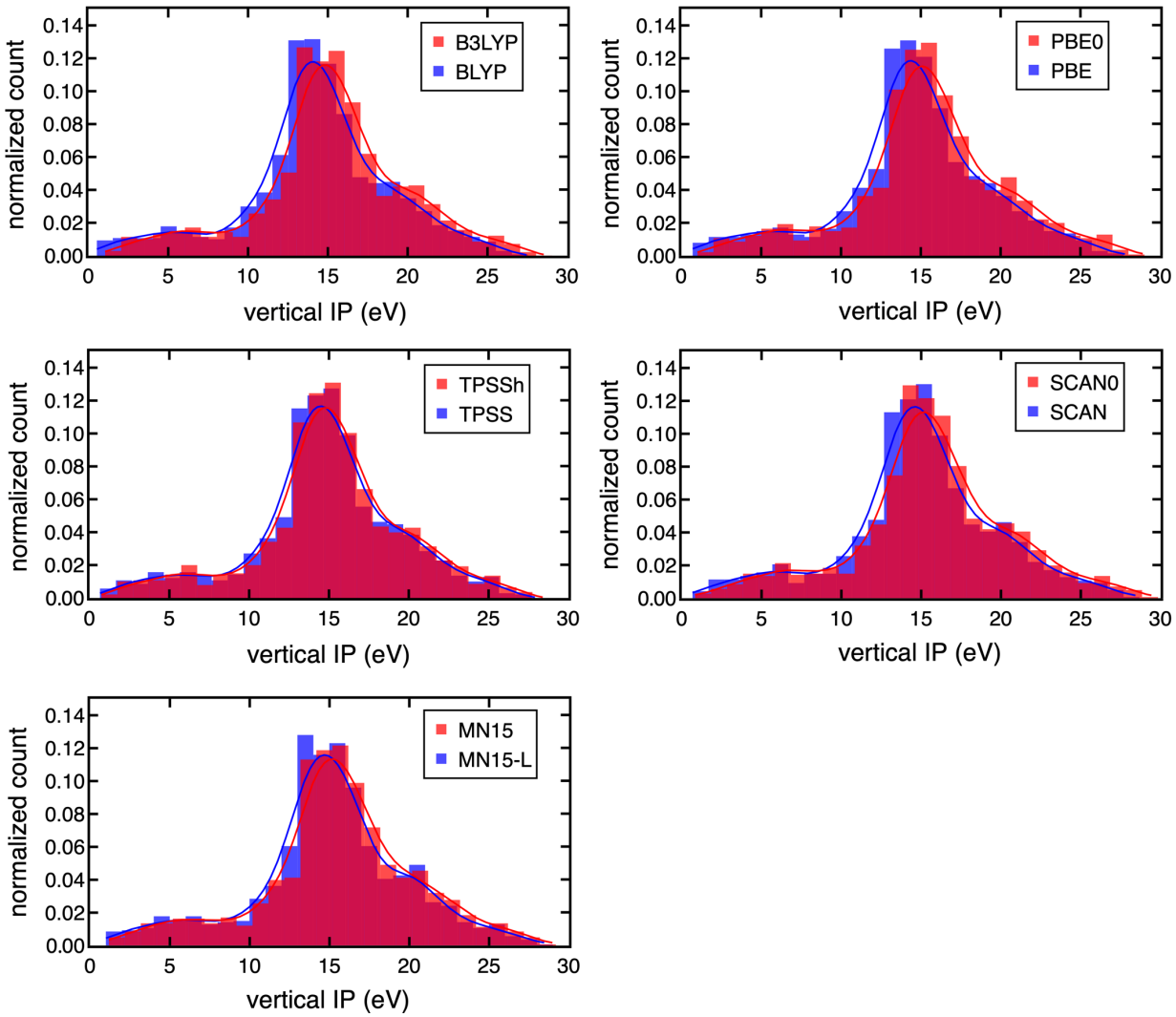

Figure S4. Distribution of vertical IP (in eV) for select DFAs (from left to right and top to bottom, as indicated in inset legend): BLYP and B3LYP (top left), PBE and PBE0 (top right), TPSS and TPSSh (middle left), SCAN and. SCAN0 (middle right), and MN15-L and MN15 (bottom left). The bin size is 1 eV for all distributions.

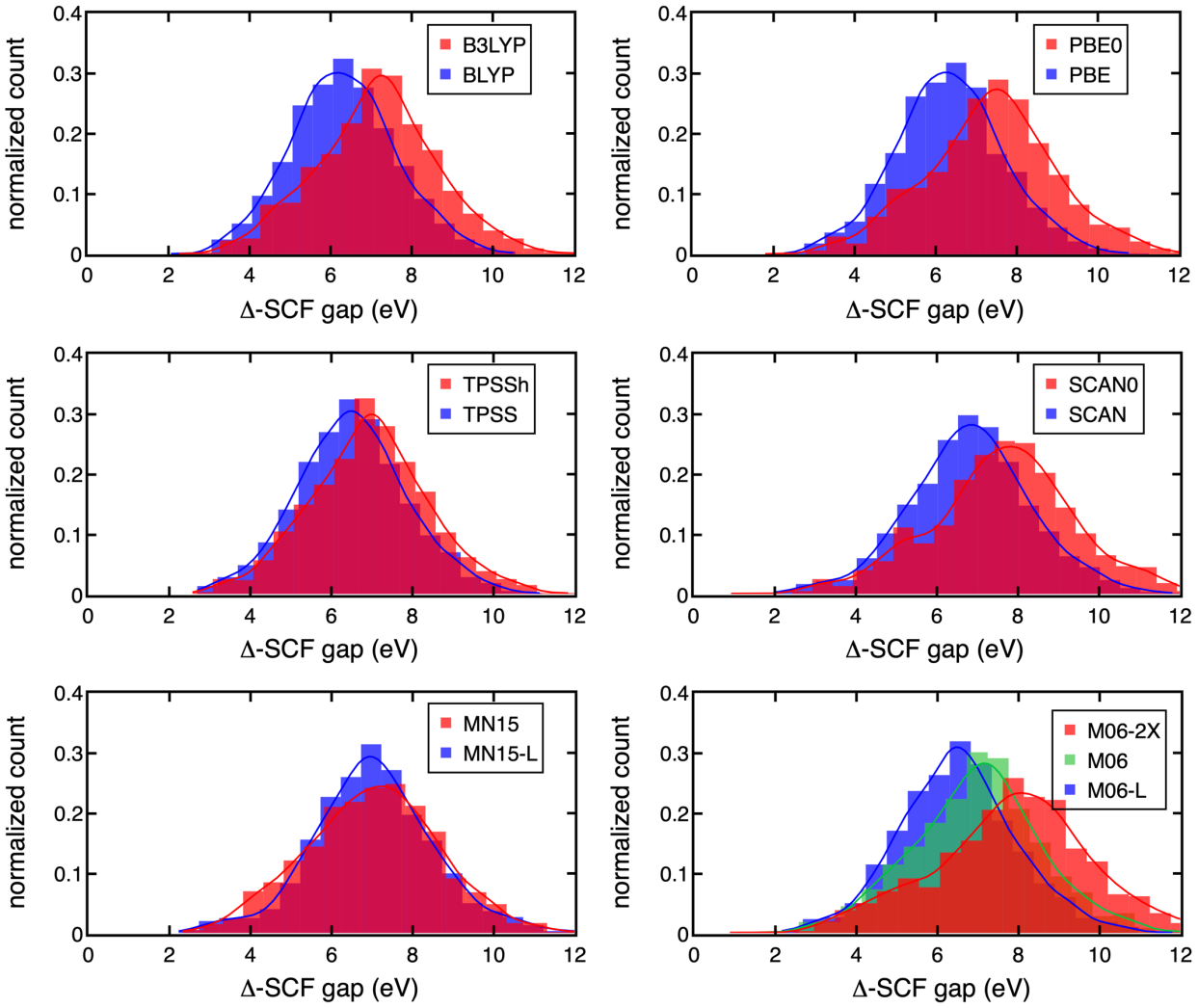

Figure S5. Distribution of Δ-SCF gap (in eV) for select DFAs (from left to right and top to bottom, as indicated in inset legend): BLYP and B3LYP (top left), PBE and PBE0 (top right), TPSS and TPSSh (middle left), SCAN and SCAN0 (middle right), MN15-L and MN15 (bottom left), and M06-L, M06, and M06-2X (bottom right). The bin size is 0.5 eV for all distributions.

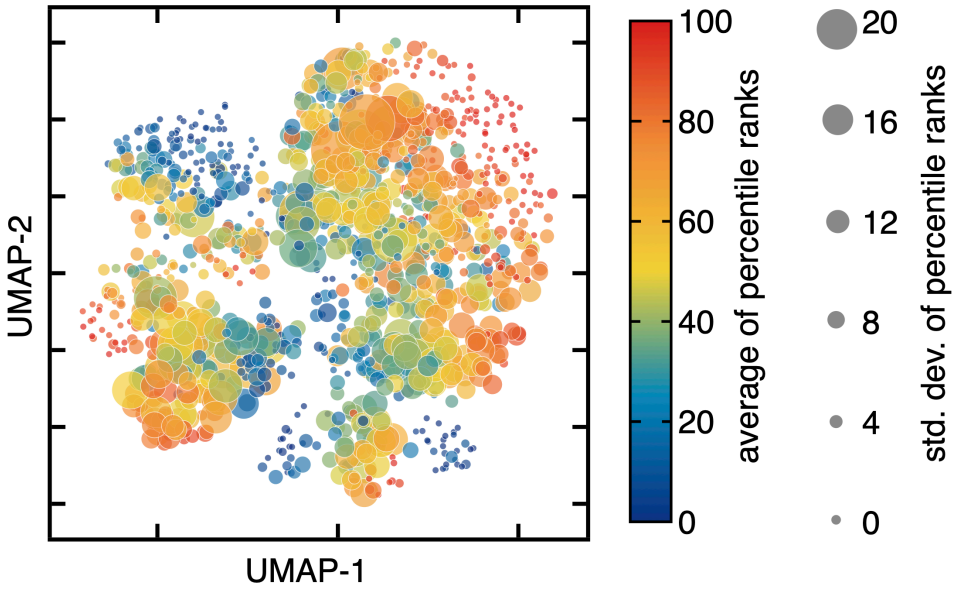

Figure S6. UMAP 2D visualization of Δ-SCF gap for the data set in the latent space of a B3LYP/LACVP* ANN trained to predict this property (see Sec. 4, main text). Each complex is represented as a circle that is colored by the average (red for high and blue for low as in inset legend) Δ-SCF gap and scaled by the std. dev. of the percentile ranks of the Δ-SCF gap from the 23 DFAs.

Table S4. Percentile ranks for each DFA and their mean value and std. dev. for five representative complexes in Figure 3 in the main text.

	$\Delta E_{H\text{-}L}$			vertical IP	
	$Co(III)(CO)_6$	$Mn(II)(H_2O)_5(pyr)$	$Mn(II)(CO)_4(H_2O)(pyr)$	$Fe(II)(NH_3)_6$	$Mn(II)(furan)_4(CO)_2$
BP86	97	3	73	48	37
BLYP	98	3	70	50	36
PBE	97	3	72	48	37
TPSS	98	3	70	46	36
SCAN	98	6	62	46	64
M06-L	98	4	67	46	35
MN15-L	97	0	45	45	35
B3LYP	98	3	50	47	59
B3P86	98	3	53	45	61
B3PW91	98	2	57	45	62
PBE0	98	2	49	44	63
ωB97X	98	3	37	44	67
LRC-ωPBEh	98	3	48	42	67
TPSSh	98	3	62	44	61
SCAN0	98	6	39	43	66
M06	98	1	50	44	64
M06-2X	97	1	20	44	66
MN15	97	0	55	33	59
B2GP-PLYP	97	0	30	44	67
PBE0-DH	98	1	34	43	67
DSD-BLYP-D3BJ	97	0	26	44	67
DSD-PBEB95-D3BJ	97	0	26	43	67
DSD-PBEP6-D3BJ	97	0	28	43	68
mean	97.6	2.3	48.8	44.3	57.0
std. dev.	0.5	1.9	16.1	3.1	12.7

Table S5. Summary of Pearson's *r* (labeled as *r*), coefficient of determination (R^2), and mean absolute difference (MAD) between a small (LACVP*) and large (def2-TZVP) basis set evaluated at each functional for three properties considered in this work.

	$\Delta E_{H\text{-}L}$ (kcal/mol)			Δ-SCF gap (eV)			vertical IP (eV)		
	r	R^2	MAD	*r*	R^2	MAD	*r*	R^2	MAD
BP86	0.99	0.97	4.2	0.99	0.95	0.24	1.00	1.00	0.15
BLYP	0.99	0.97	3.6	0.99	0.94	0.25	1.00	1.00	0.14
PBE	0.99	0.96	4.7	0.99	0.95	0.24	1.00	1.00	0.15
TPSS	0.99	0.89	8.4	0.97	0.91	0.30	1.00	0.99	0.20
SCAN	0.99	0.93	7.2	0.97	0.91	0.33	1.00	1.00	0.22
M06-L	0.97	0.88	7.4	0.97	0.93	0.24	1.00	1.00	0.18
MN15-L	0.99	0.98	3.5	0.98	0.96	0.21	1.00	1.00	0.16
B3LYP	0.98	0.95	3.5	0.96	0.91	0.31	1.00	1.00	0.16
B3P86	0.98	0.96	3.8	0.96	0.91	0.32	1.00	1.00	0.23
B3PW91	0.98	0.95	4.1	0.96	0.91	0.32	1.00	0.99	0.24
PBE0	0.98	0.95	4.2	0.96	0.91	0.33	1.00	0.99	0.25
ωB97X	0.98	0.93	4.6	0.96	0.90	0.39	1.00	0.99	0.33
LRC-ωPBEh	0.98	0.95	4.6	0.96	0.91	0.35	1.00	0.99	0.29
TPSSh	0.98	0.88	7.8	0.96	0.90	0.34	1.00	1.00	0.24
SCAN0	0.97	0.92	5.8	0.96	0.91	0.37	1.00	0.99	0.29
M06	0.97	0.93	6.2	0.95	0.91	0.32	1.00	0.99	0.22
M06-2X	0.97	0.90	6.5	0.95	0.91	0.41	1.00	0.99	0.34
MN15	0.95	0.87	8.7	0.95	0.88	0.41	1.00	0.99	0.27
B2GP-PLYP	0.97	0.90	7.1	0.96	0.87	0.50	1.00	0.99	0.23
PBE0-DH	0.98	0.93	8.0	0.96	0.91	0.40	1.00	0.99	0.27
DSD-BLYP-D3BJ	0.97	0.88	8.0	0.96	0.87	0.52	1.00	0.99	0.23
DSD-PBEB95-D3BJ	0.97	0.89	8.0	0.96	0.89	0.45	1.00	0.99	0.23
DSD-PBEP6-D3BJ	0.97	0.85	8.5	0.96	0.88	0.50	1.00	0.99	0.23

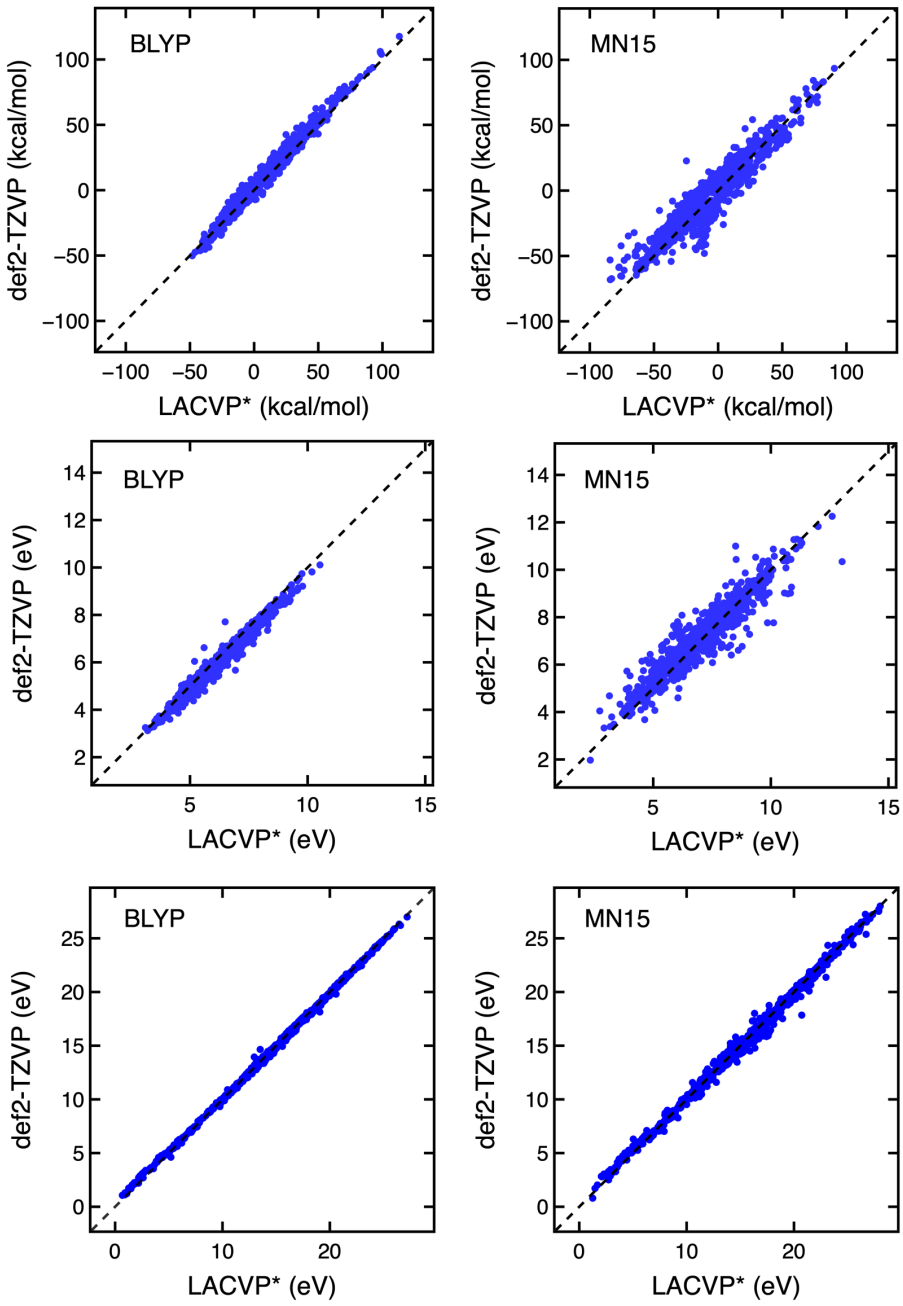

Figure S7. Parity plots of three properties, $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom), obtained with LACVP* and def2-TZVP for two representative functionals: BLYP (left) and MN15 (right).

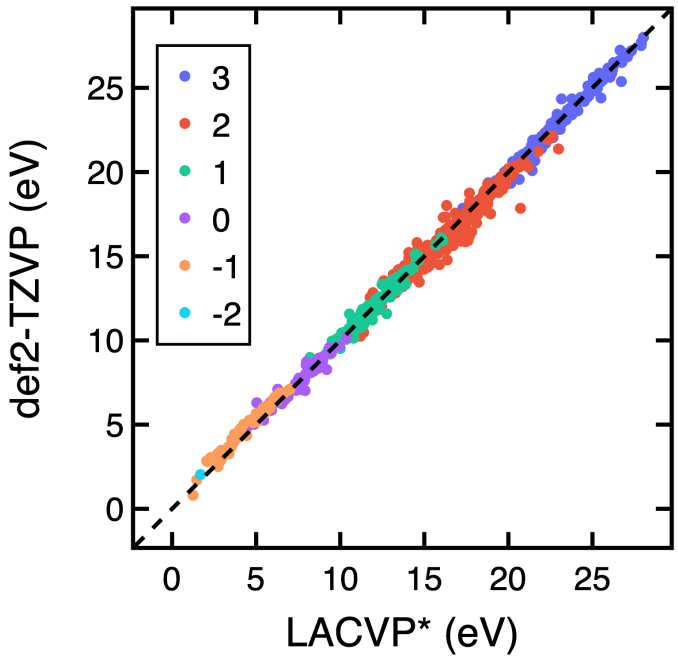

Figure S8. Parity plot of the vertical IP obtained with MN15 using LACVP* vs the def2-TZVP basis set grouped by the total charge of the TMC as indicated in inset legend.

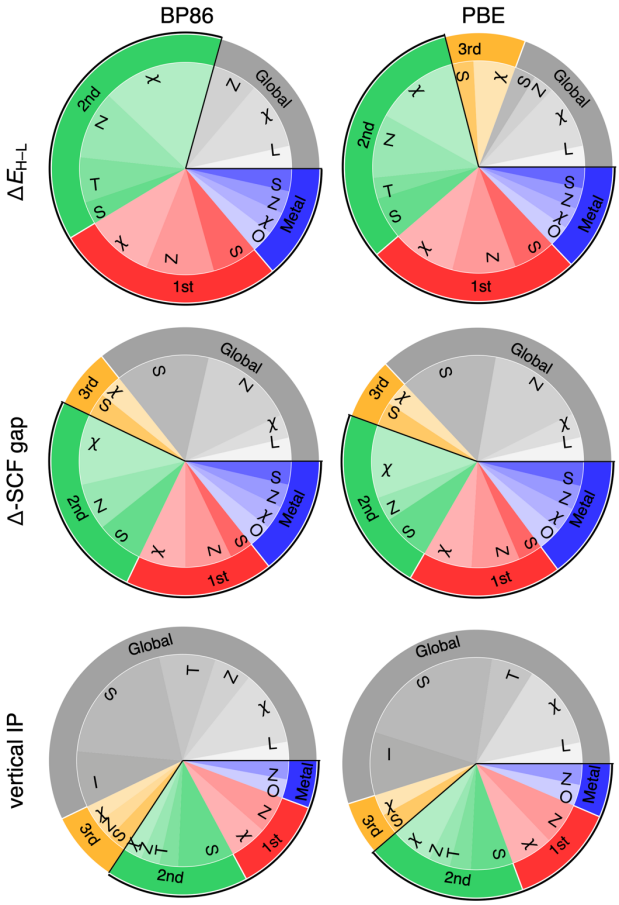

Figure S9. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for two additional GGAs that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

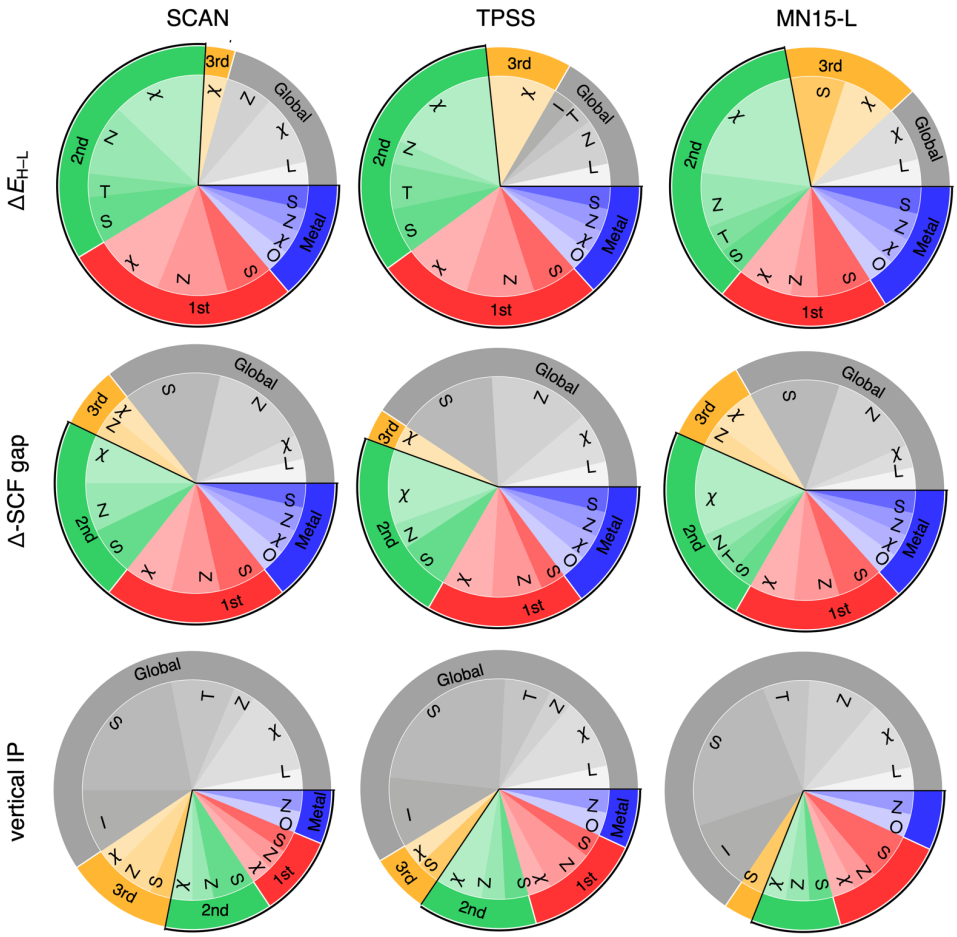

Figure S10. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for three additional meta-GGAs that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

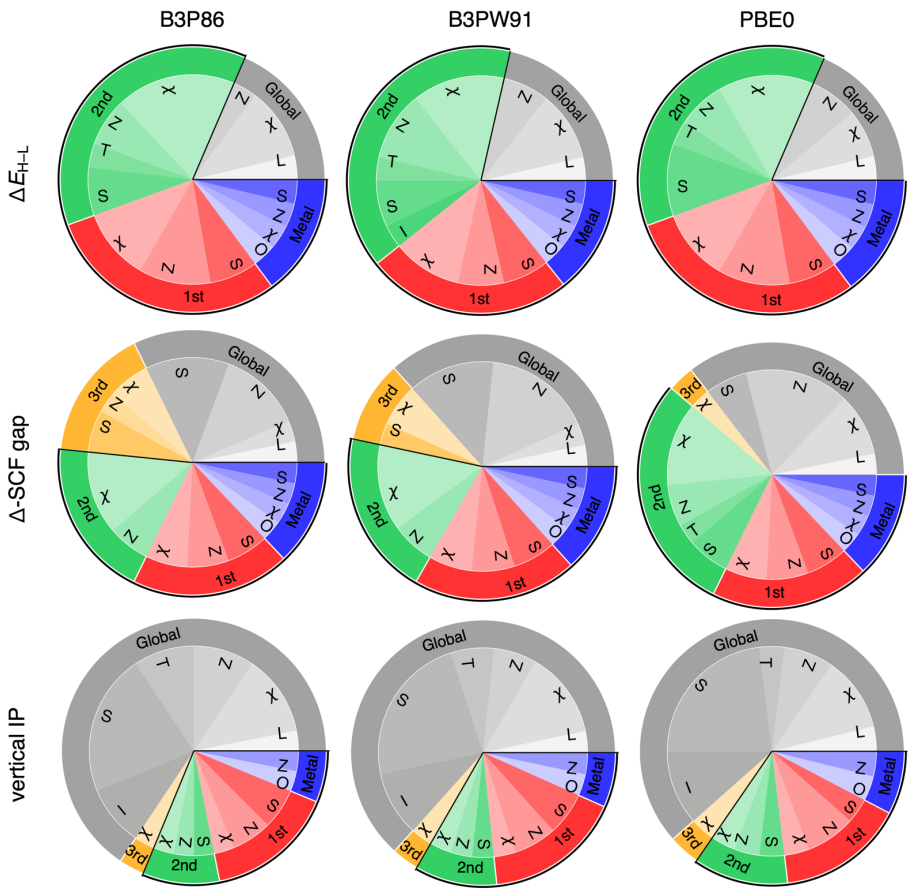

Figure S11. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for three additional GGA hybrids that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

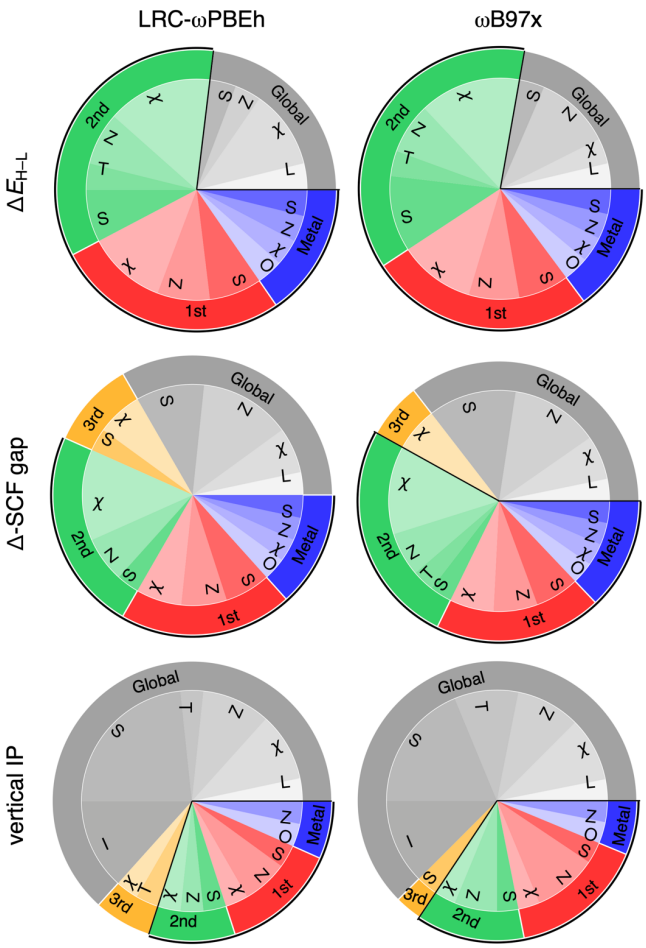

Figure S12. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for the two range-separated hybrids that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

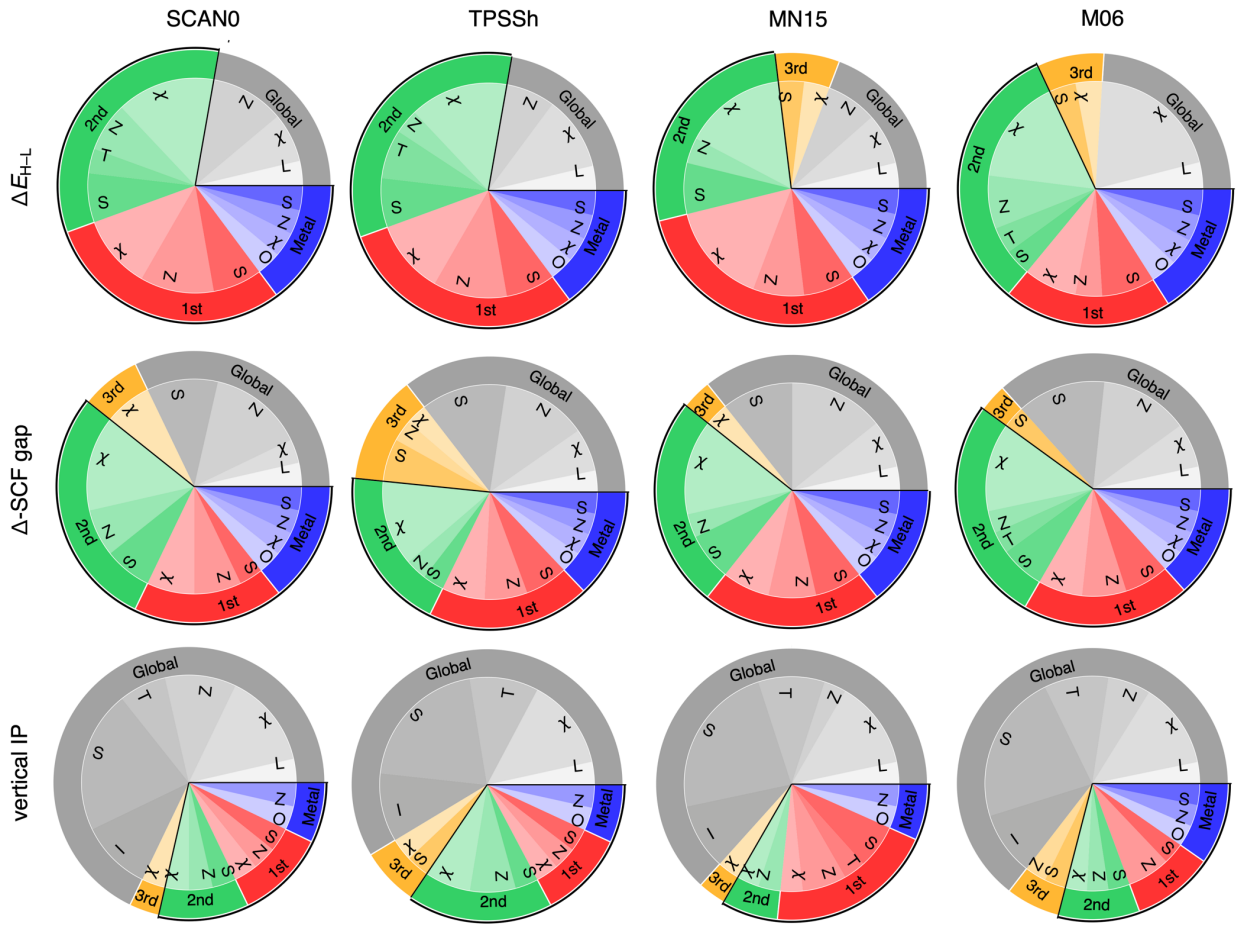

Figure S13. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for the four additional meta-GGA hybrids that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

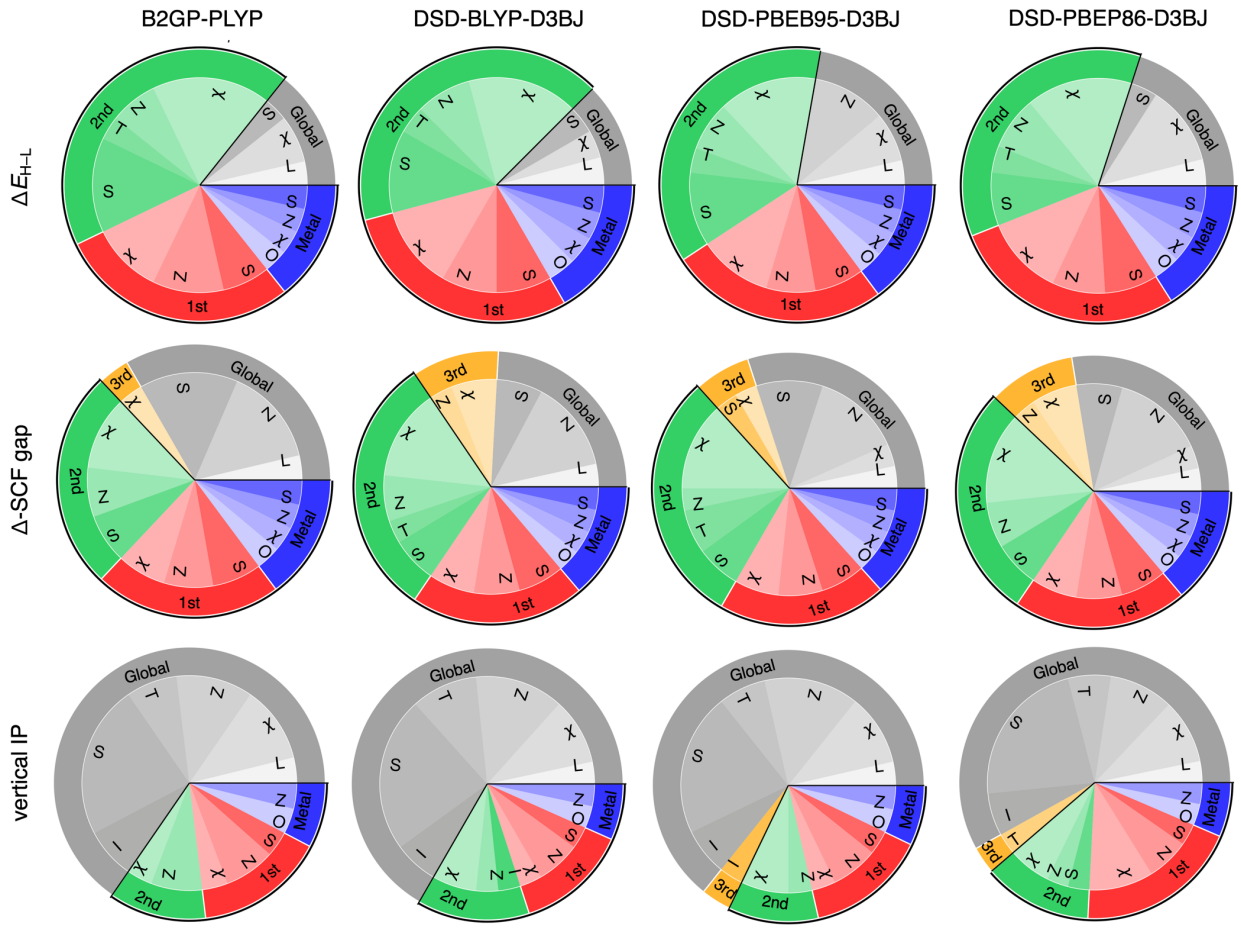

Figure S14. Pie charts of the RF-RFA/KRR-selected features for $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) for the four double hybrids that are not shown in Figure 4 in the main text. The format of the pie charts is the same as that in Figure 4.

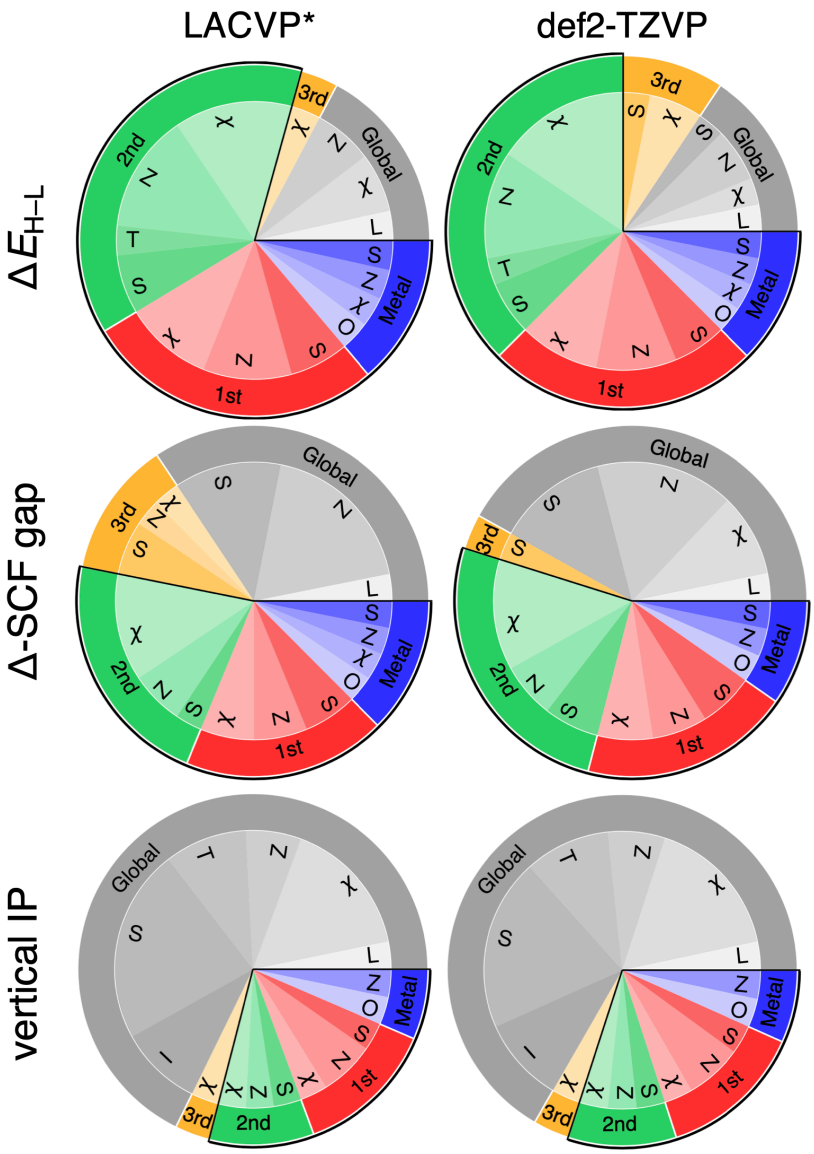

Figure S15. Pie charts of the RF-RFA/KRR-selected features selected features of RF-RFA/KRR for B3LYP $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom) with the LACVP* (left) and def2-TZVP (right) basis sets. The format of the pie charts is the same as that in Figure 4.

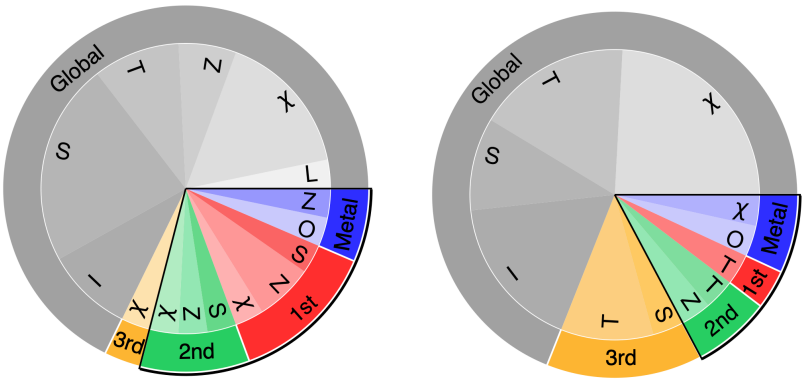

Figure S16. Pie charts of the RF-RFA/KRR-selected features for vertical IP with B3LYP/LACVP* applied to a large dataset in this work (i.e., *MD1+OHLDB*, left) and the smaller set of complexes (i.e., *SRX*) from previous work[1,2] (right).

Table S6. Hyperparameters for fine-tuned ANN (FT-ANN) models obtained for each property and basis set combination. To obtain these models, we re-optimized the model weights of a B3LYP ANN model with a reduced (i.e., 1e-5) learning rate for each of the 23 functionals at each basis set and property combination. The hyperparameters of the FT-ANN models at each basis set and property combination are the same as those in the original B3LYP ANN model.

	LACVP*			def2-TZVP		
	$\Delta E_{H\text{-}L}$	vertical IP	Δ-SCF gap	$\Delta E_{H\text{-}L}$	vertical IP	Δ-SCF gap
Architecture	[512,512, 512]	[256,256,256]	[512,512, 512]	[256,256,256]	[256,256,256]	[256,256,256]
L2 regularization	2.2e-4	1.0e-4	4.8e-4	2.0e-3	3.0e-3	5.6e-2
Dropout rate	0.41	0.30	0.02	0.08	0.25	0.28
Learning rate	3.3e-4	7.3e-4	4.8e-4	7.5e-4	9.0e-4	8.8e-4
Beta1	0.88	0.85	0.88	0.94	0.88	0.89
Batch size	128	256	32	256	128	32

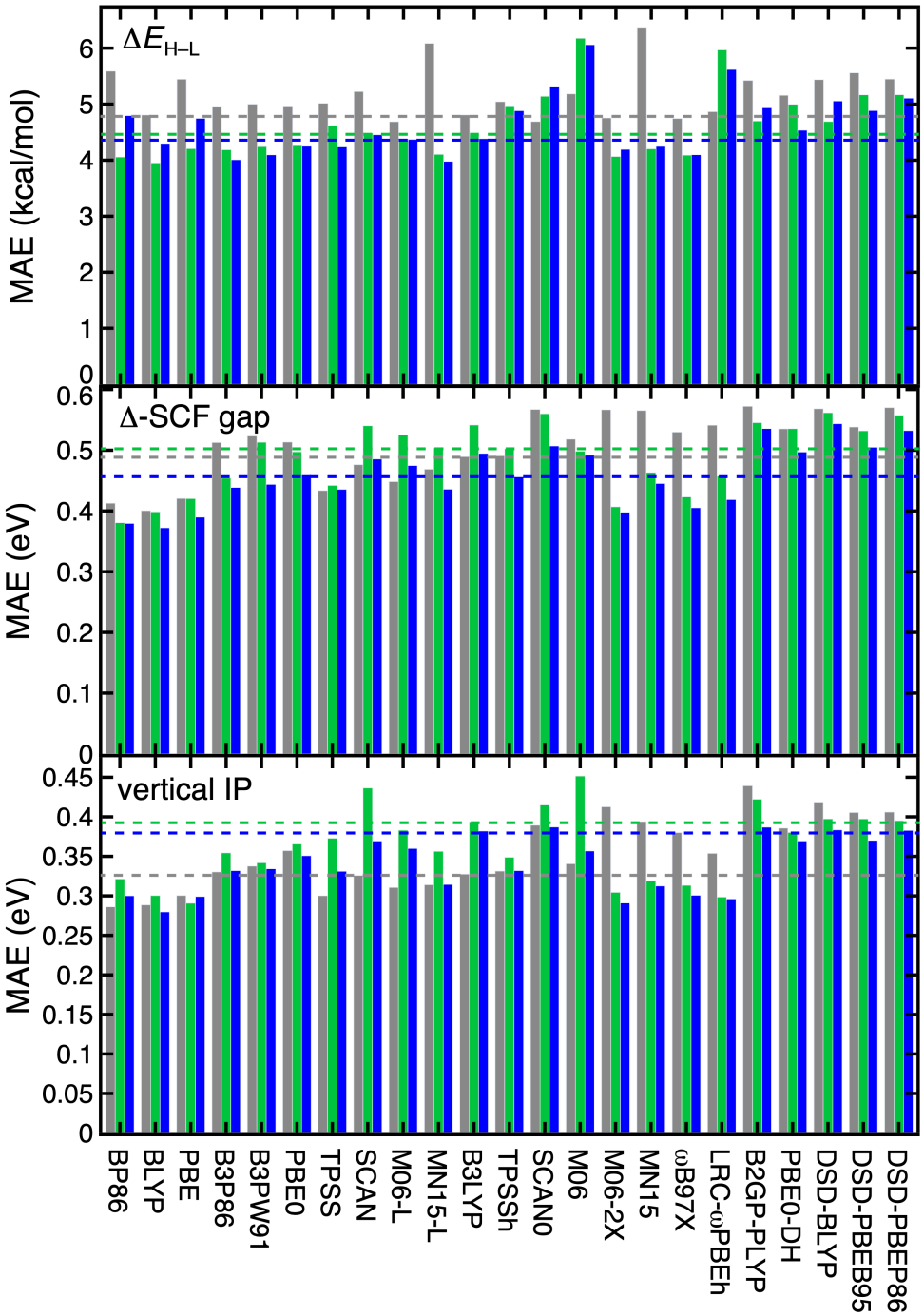

Figure S17. Mean absolute error (MAE) for three types of models: RF-RFA/KRR (gray), ANN (green), and FT-ANN (blue) of all 23 DFAs with the LACVP* basis set for three properties: $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom). A horizontal dashed line is shown for the B3LYP MAE for each type of model as a reference. The D3BJ dispersion correction is included in all three DSD double hybrids.

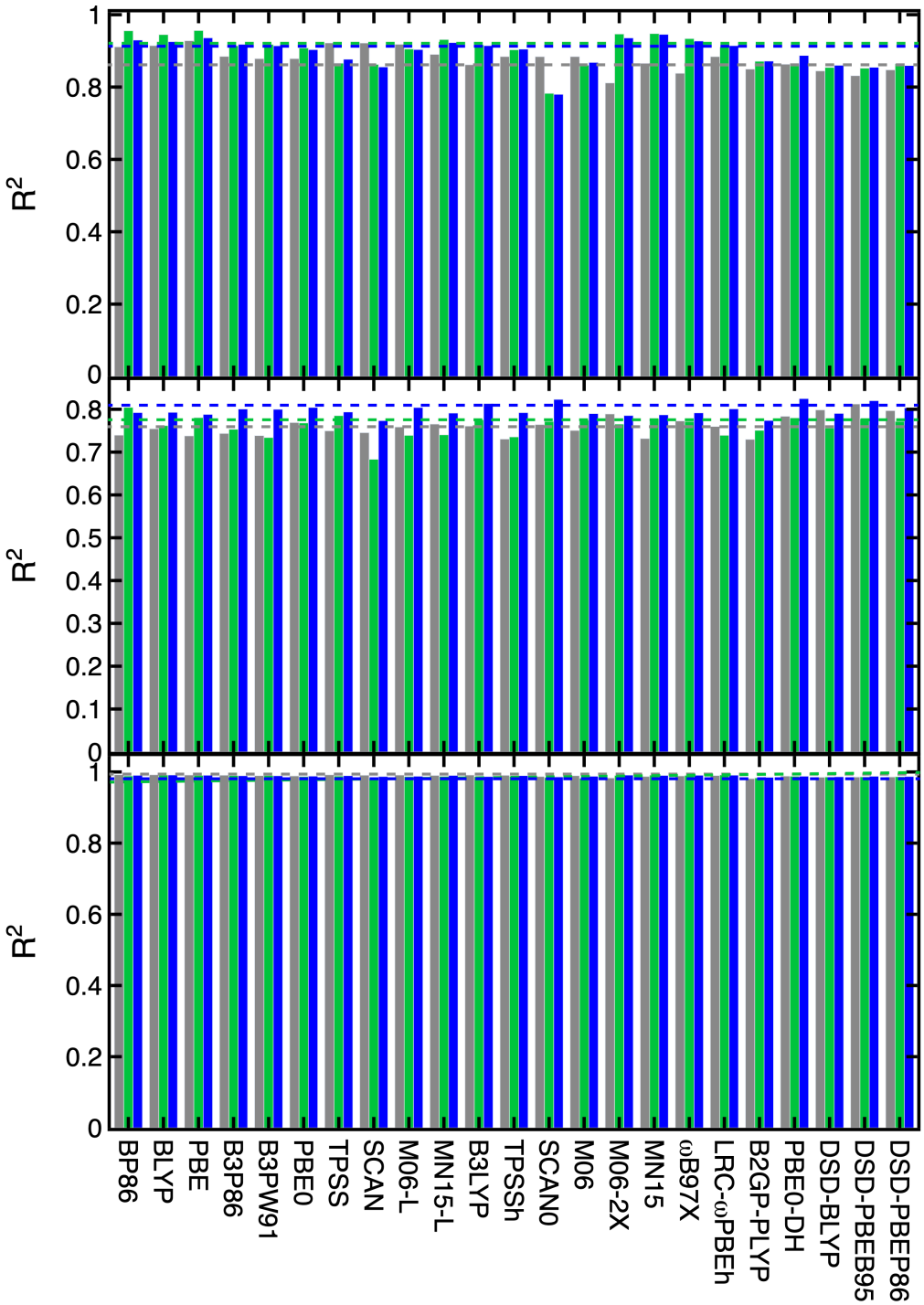

Figure S18. Coefficient of determination (R^2) for three types of models: RF-RFA/KRR (gray), ANN (green), and FT-ANN (blue) of all 23 DFAs with the LACVP* basis set for three properties: $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom). A horizontal dashed line is shown for B3LYP R^2 for each type of model as a reference. The D3BJ dispersion correction is applied in all three DSD double hybrids.

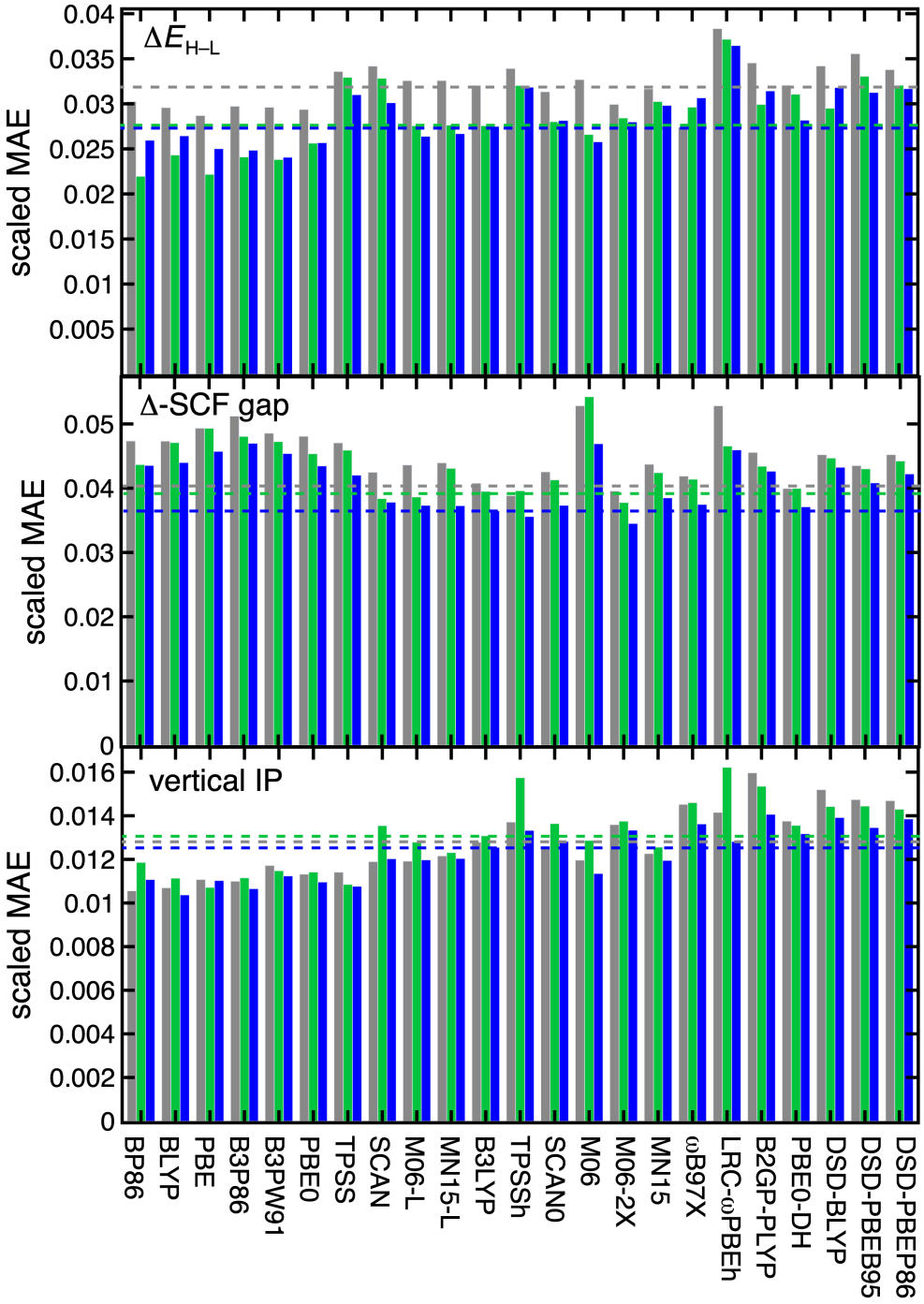

Figure S19. Scaled MAE (i.e., by the training set property range) for three types of models: RF-RFA/KRR (gray), ANN (green), and FT-ANN (blue) of all 23 DFAs with the LACVP* basis set for three properties: $\Delta E_{\text{H-L}}$ (top), Δ-SCF gap (middle), and vertical IP (bottom). A horizontal dashed line is shown for the B3LYP scaled MAE for reference. The D3BJ dispersion correction is applied in all three DSD double hybrids.

Table S7. Average value (mean) and standard deviation (std. dev.) of the MAE of model performance of 23 functionals with the LACVP* basis set.

	$\Delta E_{H\text{-}L}$ (kcal/mol)			Δ-SCF gap (eV)			vertical IP (eV)		
	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN
mean	5.2	4.6	4.6	0.51	0.49	0.46	0.35	0.36	0.34
std. dev	0.4	0.6	0.5	0.05	0.05	0.05	0.04	0.05	0.03

Table S8. Average value (mean) and standard deviation (std. dev.) of the R^2 of model performance of 23 functionals with the LACVP* basis set.

	$\Delta E_{H\text{-}L}$			Δ-SCF gap			vertical IP		
	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN
mean	0.88	0.90	0.90	0.76	0.76	0.79	0.99	0.99	0.99
std. dev	0.03	0.04	0.04	0.02	0.02	0.01	0.00	0.00	0.00

Table S9. Average value (mean) and standard deviation (std. dev.) of the scaled MAE of model performance of 23 functionals with the LACVP* basis set.

	$\Delta E_{H\text{-}L}$			Δ-SCF gap			vertical IP		
	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN	RF-RFA KRR	ANN	FT-ANN
mean	0.032	0.029	0.029	0.045	0.044	0.041	0.013	0.013	0.012
std. dev	0.002	0.004	0.003	0.004	0.004	0.004	0.002	0.002	0.001

Table S10. Categories of DFAs studied in this work on each rung of "Jacob's ladder" or by inclusions of HF exchange.

semi-local (7)		hybrid (11)			double hybrid (5)
GGA	meta-GGA	GGA hybrid	meta-GGA hybrid	range-separated hybrid	double hybrid
3	4	4	5	2	5

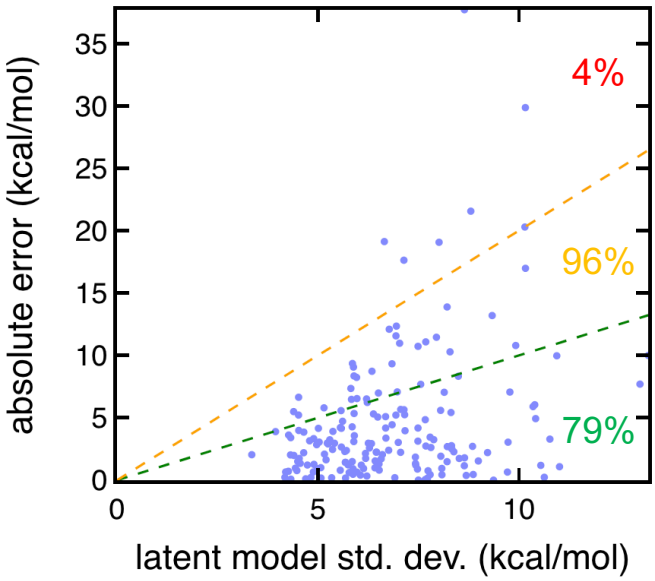

Figure S20. Spin-splitting ANN model trained on B3LYP/LACVP* data with absolute errors vs the uncertainty metric[3] (labeled as latent model std. dev.,), with both in kcal/mol. The model uncertainty is calibrated against the set-aside test data with maximum likelihood following our established procedure[3]. The green and yellow dashed lines correspond to one and two std. dev., respectively. 79% of the complexes are within one std. dev. and 96% of the complexes are within two std. dev.

Table S11. Cutoffs for latent model std. dev. (i.e., calibrated distance in latent space) used during the exploration of SCO complexes ($|\Delta E_{H\text{-}L}| < 5$ kcal/mol) and complexes at a targeted Δ-SCF gap (i.e., < 3 eV). These cutoffs are applied to the hypothetical space of 187.2k complexes. In practice, this cutoff leads to making predictions on only about 67% of the hypothetical space.

SCO complexes	20 kcal/mol
Targeted-gap complexes	2 eV

Table S12. Design space ligands with the net charge (charge), denticity (dent), number of atoms (n_{at}), connecting atom type (CA), and SMILES string. These ligands are the unique ligands in the dataset from "merged dataset", *MD1*, introduced in Ref. 4.

id	name	charge	dent.	n_{at}	CA	SMILES
1	acac	-1	2	14	O	O=C(C)C=[CH-](O)C
2	aceticacidbipyridine	0	2	32	N	n1ccc(cc1c1nccc(c1)CC(=O)O)CC(=O)O
3	acetonitrile	0	1	6	N	N#CC
4	ammonia	0	1	4	N	N
5	benzisc	0	1	16	C	[C-]#[N+]Cc1ccccc
6	bipy	0	2	20	O	n1ccccc1c1ncccc1
7	carbonyl	0	1	2	N	C#[O]
8	cyanide	-1	1	2	C	[C-]#N
9	cyanoaceticporphyrin	-2	4	52	N	N1=C2C=[CH2-][CH-]1=C(c1[nH]c(cc1)/C=C/1\N=C(/C(=c/3\[nH]/c(=C\2)/cc3)/C=C(/C(=O)O)\C#N)C=C1)/C=C(/C(=O)O)\C#N
10	cyanopyridine	0	1	12	N	C1(=CCNC=C1)C#N
11	en	0	2	12	N	NCCN
12	formaldehyde	0	1	4	O	C=O
13	furan	0	1	9	O	o1cccc1
14	isothiocyanate	-1	1	3	N	[N-]=C=S
15	mebipyridine	0	2	26	N	n1ccc(cc1c1nccc(c1)C)C
16	mec	-2	2	15	O	[O-]c1c(cc(cc1)C)[O-]
17	methylamine	0	1	7	N	CN
18	misc	0	1	6	C	[C-]#[N+]C
19	ox	-2	2	6	O	C(=O)(C(=O)[O-])[O-]
20	phen	0	2	22	N	c1cc2ccc3cccnc3c2nc1
21	phenisc	0	1	13	C	[C-]#[N+]c1ccccc1
22	pisc	0	1	25	C	[C-]#[N+]c1ccc(C(C)(C)C)cc1
23	porphyrin	-2	4	36	N	N1=C2C=[CH2-][CH-]1=Cc1[nH]c(cc1)/C=C/1\N=C(/C=c/3\[nH]/c(=C\2)/cc3)C=C1
24	pph3	0	1	34	P	c1c(P(c2ccccc2)c2ccccc2)cccc1
25	py	0	1	11	C	C1=CCNC=C1
26	tbuc	-2	2	24	O	[O-]c1c(cc(C(C)(C)C)cc1)[O-]
27	thiopyridine	0	1	12	N	C1(=CCNC=C1)S
28	water	0	1	3	O	O
29	fluoride	-1	1	1	F	[F-]
30	iodide	-1	1	1	I	[I-]
31	[O-][O-]	-2	1	2	O	[O-][O-]
32	hydroxyl	-1	1	2	O	[OH-]
33	phosphine	0	1	4	P	[PH3]
34	[S--]	-2	1	1	S	[S--]
35	hydrogen sulfide	0	1	3	S	[SH2]
36	cyanate	-1	1	3	N	N#C[O-]

Table S13. Allowed ligand combinations in the theoretical complex space according to the symmetry and allowed equatorial or axial ligand type. The 11,700 complexes are combined with 16 possible metal/oxidation/spin state combinations to produce a theoretical space of 187,200 complexes.

class	Allowed eq	Allowed ax	total
homoleptic	monodentate (25)	--	25
heteroleptic, ax1 = ax2	monodentate (25)	monodentate different from eq (24)	25×24 = 600
	bidentate (9)	monodentate (25)	9×25 = 225
	tetradentate (2)	monodentate (25)	2×25 = 50
heteroleptic, ax1 ≠ ax2	monodentate (25)	two monodentate ligands $\binom{25}{2}$	25×300 = 7500
	bidentate (9)		9×300 = 2700
	tetradentate (2)		2×300 = 600
Total			11700

Table S14. Definitions of spin multiplicities ($2S+1$) for each metal and oxidation state in the theoretical complex space.

		M(II) multiplicity	M(III) multiplicity
Cr	LS	1	2
	HS	5	4
Mn	LS	2	1
	HS	6	5
Fe	LS	1	2
	HS	5	6
Co	LS	2	1
	HS	4	5

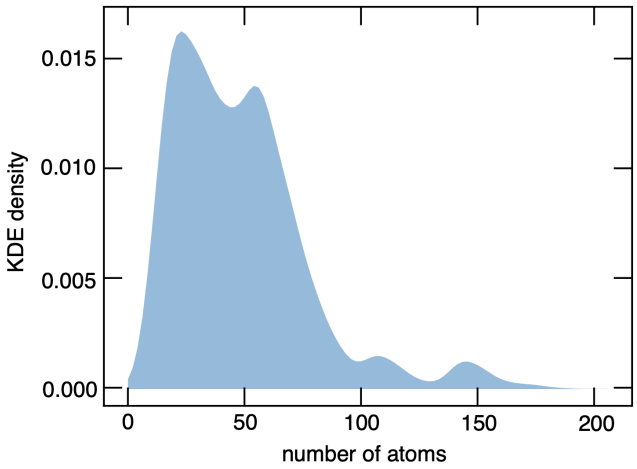

Figure S21. Kernel density estimation (KDE) of the size distribution of the 187,200 complexes design space.

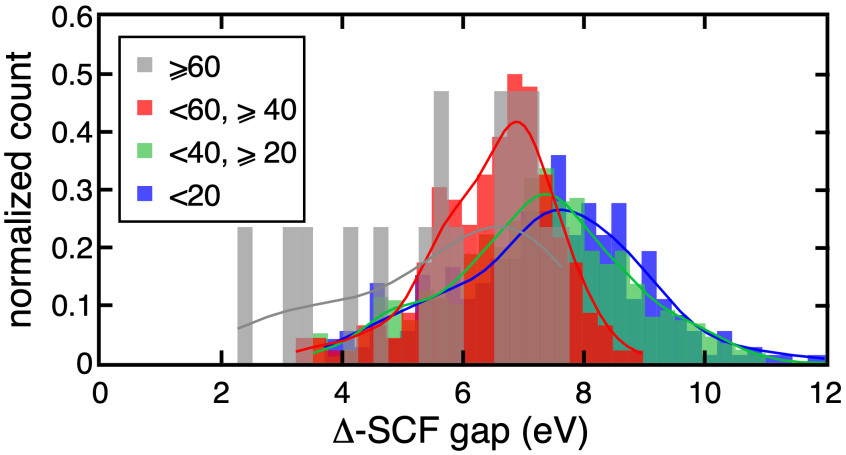

Figure S22. Normalized histograms of Δ-SCF gap over the B3LYP/LACVP* data grouped by the system size: < 20 (blue), between 20 and 40 (green), between 40 and 60 (red), and > 60 (gray).

Table S15. Density functional categorization used in Figure 8 in the main text.

semi-local (GGA and meta-GGA)	BLYP, BP86, PBE, SCAN, TPSS, M06-L, MN15-L
hybrid (range-separated and global)	B3LYP, B3P86, B3PW91, PBE0, SCAN0, TPSSh, M06, M06-2X, MN15, LRC-ωPBEh, ωB97x
double hybrid	B2GP-PLYP, PBE0-DH, DSD-BLYP-D3BJ, DSD-PBEB95-D3BJ, DSD-PBEP86-D3BJ

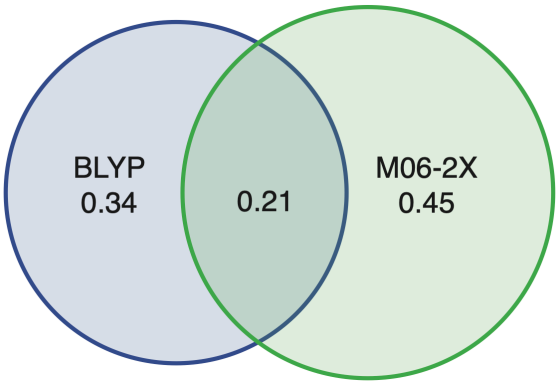

Figure S23. Venn diagram of lead targeted gap (Δ-SCF gap < 3 eV) complexes discovered by BLYP and M06-2X.

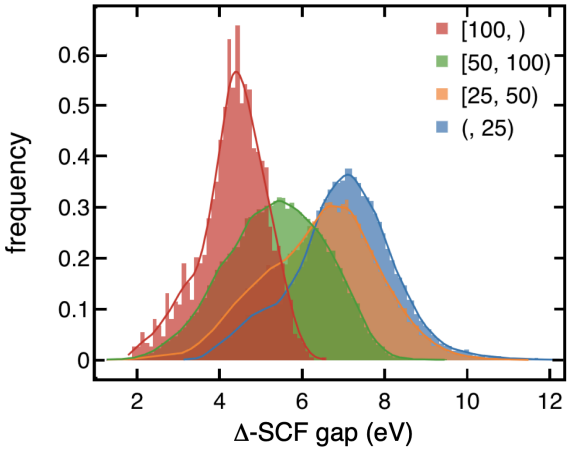

Figure S24. Distribution of predicted Δ-SCF gap from the B3LYP ANN model with different complex size ranges in the 187,200 complexes design space: below 25 atoms (blue), between 25 and 50 atoms (orange), between 50 and 100 atoms (green), above 100 atoms (red).

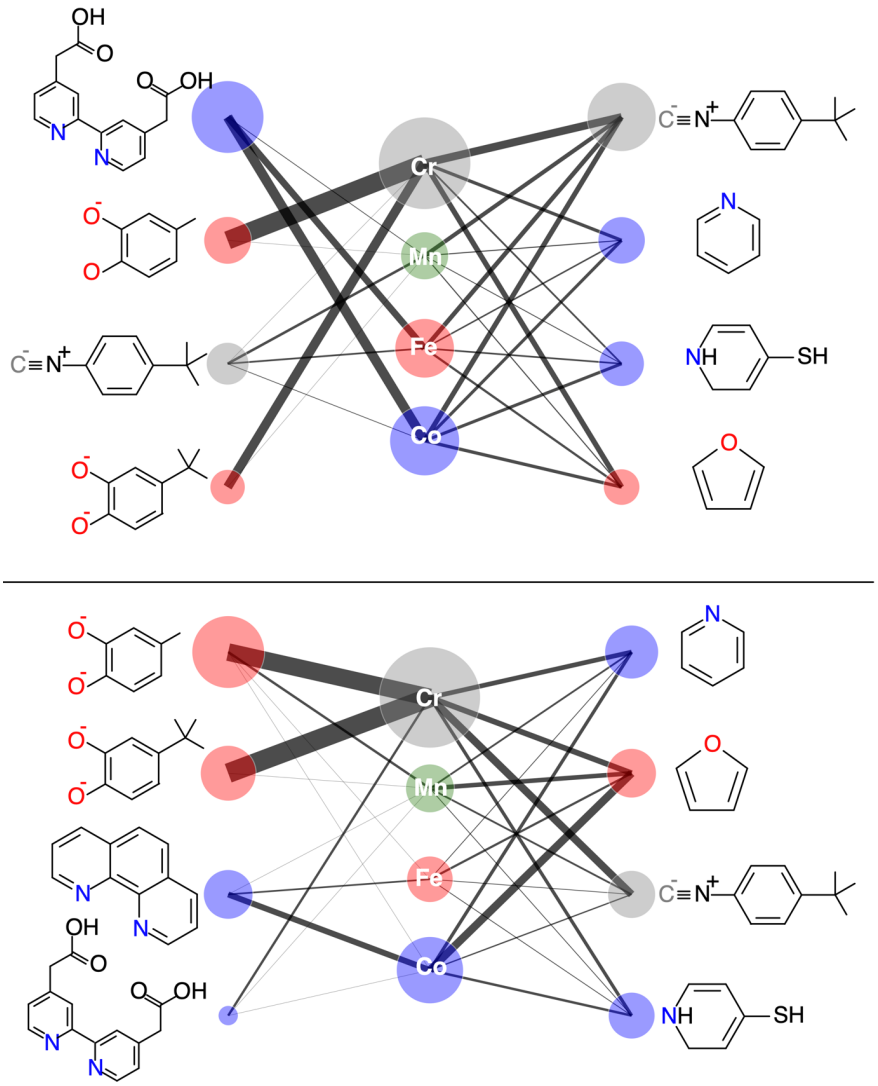

Figure S25. Network graph of the statistics for lead targeted Δ-SCF gap complexes from the BLYP FT-ANN (top) and B3LYP ANN (bottom). The scale of the circle indicates the relative abundance of the metal or equatorial/axial ligand appearing in the leads, and the width of a line connecting a metal and a ligand shows the relative abundance of this metal–ligand combination appearing in the leads. Metals are colored as the following: gray for Cr, green for Mn, red for Fe, and blue for Co. Coordinating atom types are colored as the following: gray for C, blue for N, and red for O.

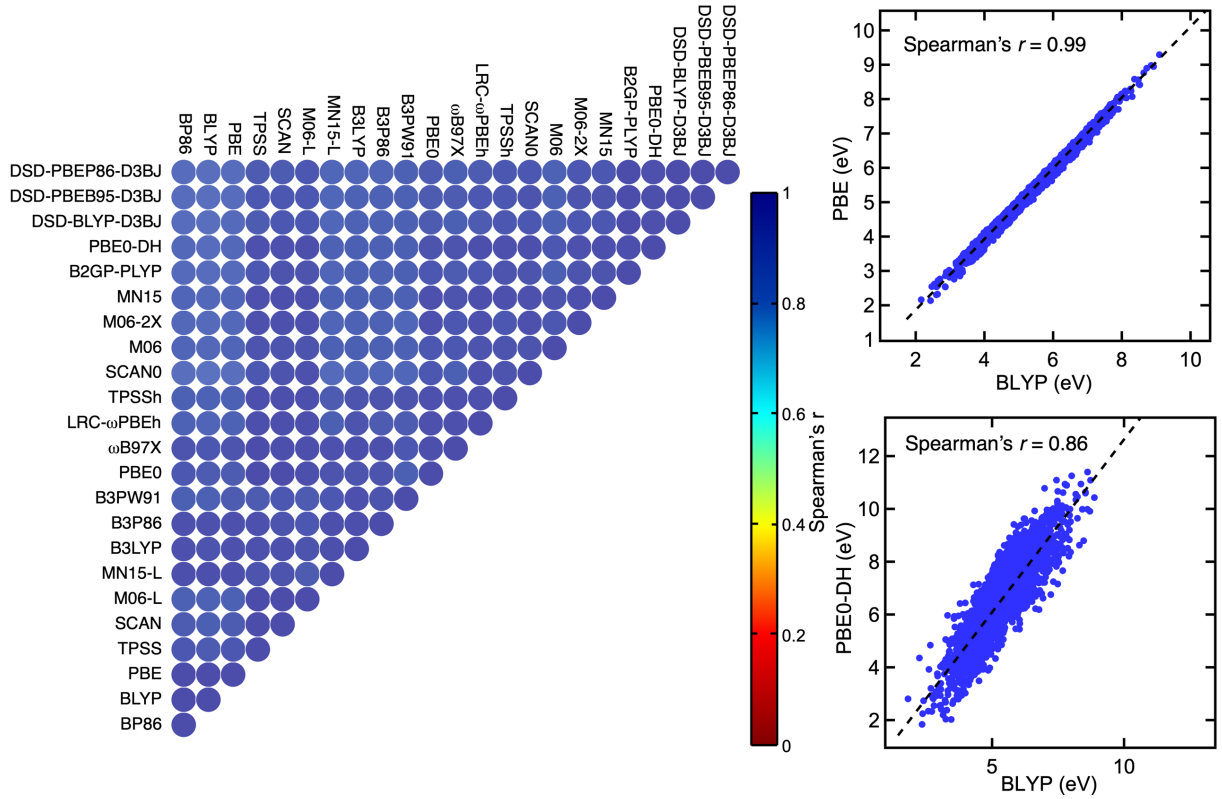

Figure S26. An upper triangular matrix colored by Spearman's *r* for pairs of 23 functionals for predicted Δ-SCF gap on the 187,200 TMCs design space (shown at left) along with a parity plot of Δ-SCF gap for PBE and BLYP (top right, highest Spearman's *r*, 0.99) and for PBE0-DH and BLYP (bottom right, lowest Spearman's *r*, 0.86).

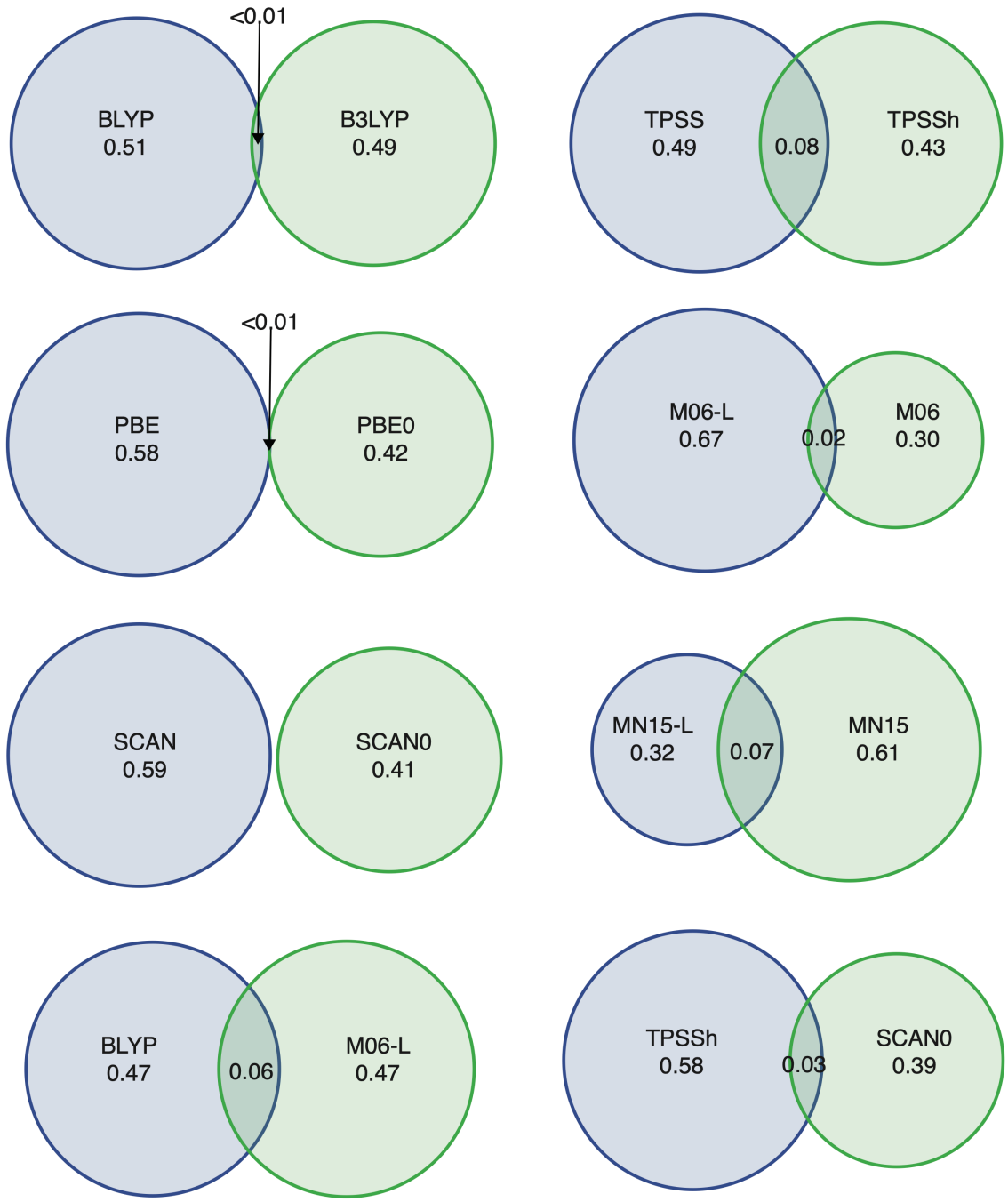

Figure S27. Venn diagrams of lead SCO complexes discovered by different pairs of DFAs as indicated.

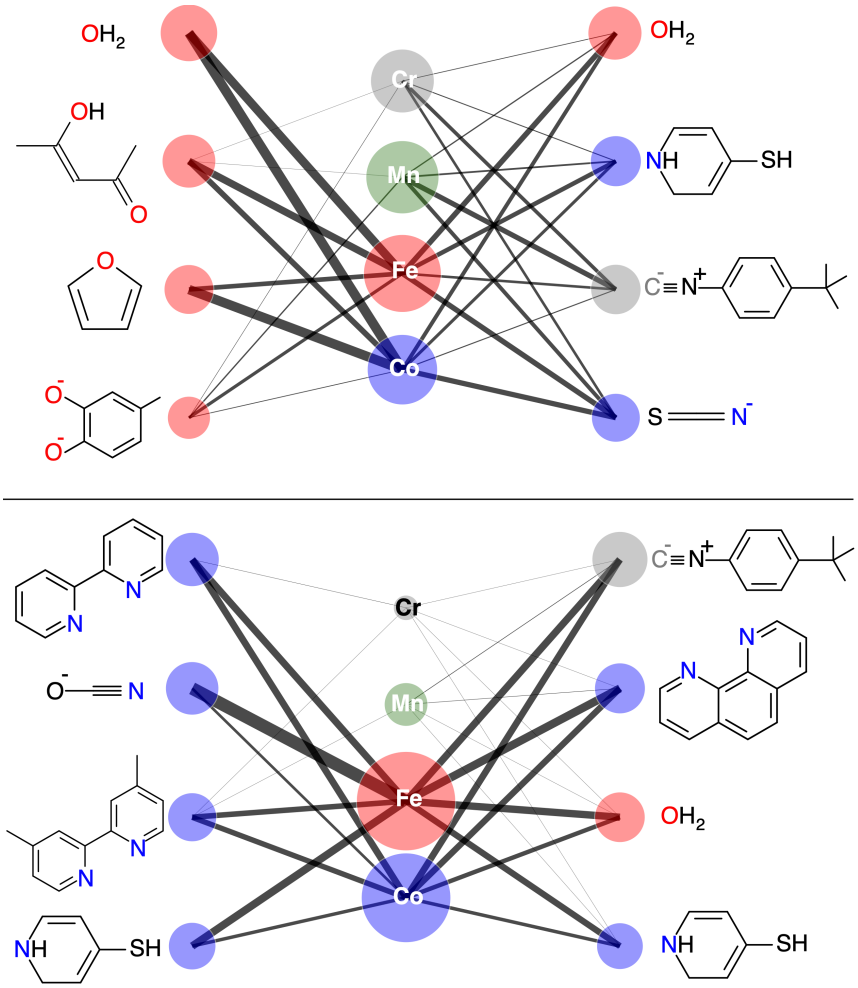

Figure S28. Network graph of the statistics of lead SCO complexes from the BLYP FT-ANN (top) and B3LYP ANN (bottom). The scale of the circle indicates the relative abundance of the metal or equatorial/axial ligand appearing in the leads, and the width of a line connecting a metal and a ligand shows the relative abundance of this metal–ligand combination appearing in the leads. Metals are colored as the following: gray for Cr, green for Mn, red for Fe, and blue for Co. Coordinating atom types are colored as the following: gray for C, blue for N, and red for O.

Text S1. Extraction of candidate SCO complexes from the literature.

Identification of candidate experimental spin-crossover (SCO) complexes was performed similarly to previous automated mining of Fe(II) SCOs.[5] Briefly, a search through the Cambridge Structural Database (CSD version 5.41(Nov. 2019) + 3 Data updates) was performed for all octahedral mononuclear transition-metal (M = Cr, Mn, Fe, Co) complexes (n=29,540).[5,6] From this set of complexes, structures with identical 6-letter refcodes but taken at different temperatures were selected. Multiple temperatures for identical structures are frequently reported in tests for spin-crossover behavior (n=5,093 structures from 1,768 6-letter refcodes).[5] For the set of identical structures from each refcode the user-labelled oxidation states were verified to be 2 or 3 and the lowest-temperature and highest-temperature structures were identified (n=1934 structures from 967 6-letter refcodes). For each of the lowest-temperature structures the abstracts and titles were mined using pybliometrics[7] and SCO keywords and sentiment were analyzed.[5] Structures that contained SCO keywords and positive sentiment were retained resulting in 452 pairs of structures. From this set of likely SCO complexes we removed structures in which at least one of the structures was identified as having user-assigned charges and where the low-temperature and high-temperature structures were from the same paper resulting in 279 pairs of structures we refer to as candidate experimental SCO complexes.

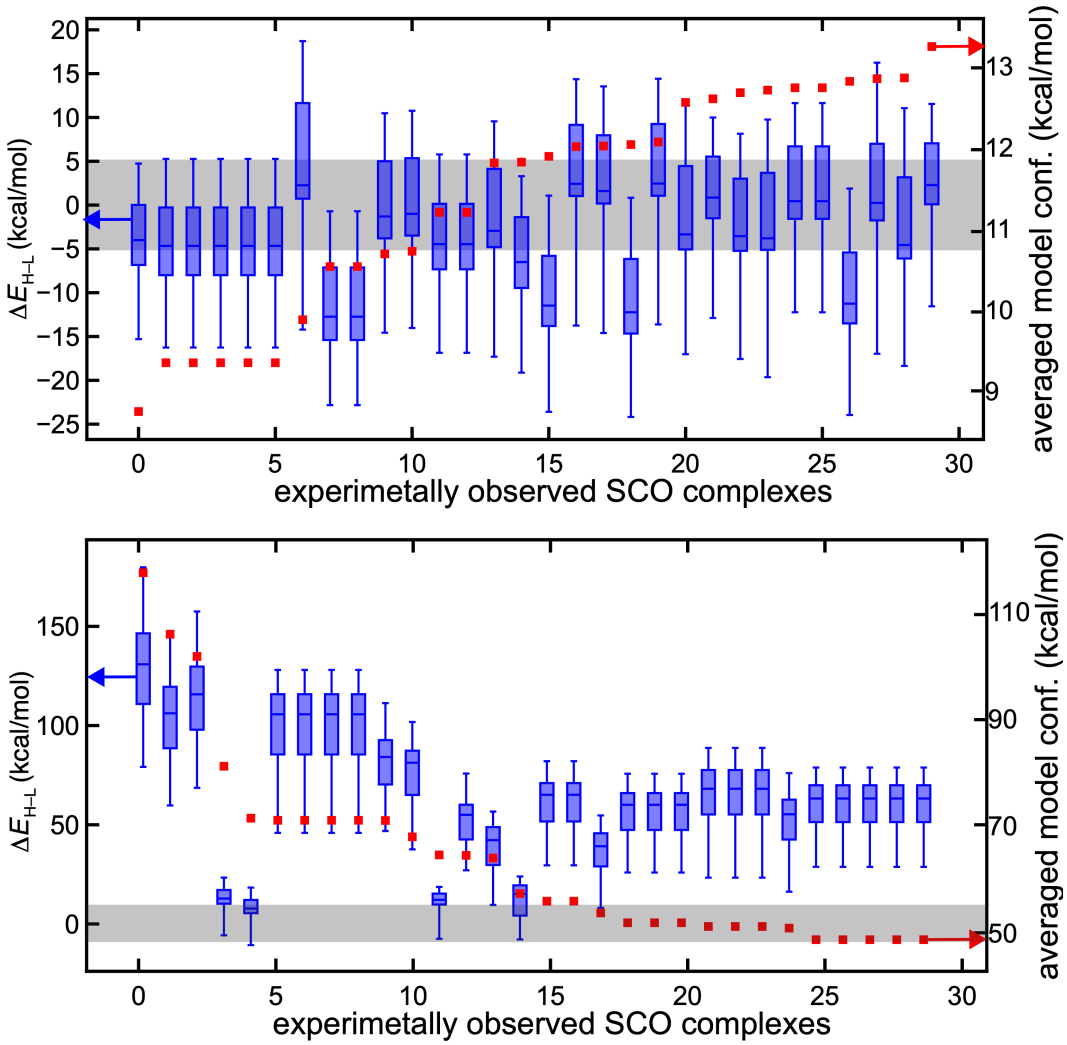

Figure S29. Box plot of $\Delta E_{H\text{-}L}$ (blue, left y axis) and averaged model confidence (red, right y axis) for the 30 most confident (i.e., top 10%) complexes from the set of experimentally observed SCO complexes (top) and 30 least confident (i.e., bottom 10%) complexes from the set of experimentally observed SCO complexes (bottom). Each box at a complex displays the median, first quarter, third quarter, minimum, and maximum value of $\Delta E_{H\text{-}L}$ predicted by 23 FT-ANNs that were trained on 23 DFAs. In each subplot, the complexes are ordered by the averaged model confidence. The shaded area corresponds to the $\Delta E_{H\text{-}L}$ for SCO complexes (i.e., $|\Delta E_{H\text{-}L}| <$ 5 kcal/mol).

Table S16. Identified experimental SCO complex counts by metal and oxidation state.

metal	ox	count
Co	2	30
	3	1
Cr	3	1
Fe	2	153
	3	70
Mn	3	24

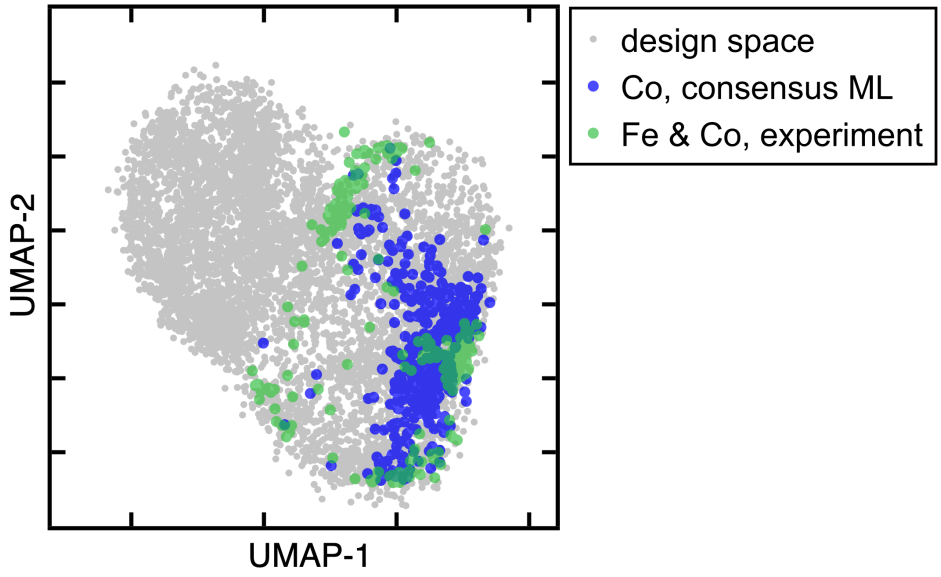

Figure S30. UMAP 2D visualization (in the latent space of the B3LYP/LACVP* FT-ANN model, see Sec. 4) of $\Delta E_{\text{H-L}}$ for the design space of 187,200 complexes (gray), lead Co SCO complexes discovered with the consensus of all 23 DFAs considered in this work (blue), and experimentally observed Fe and Co SCO complexes (green).

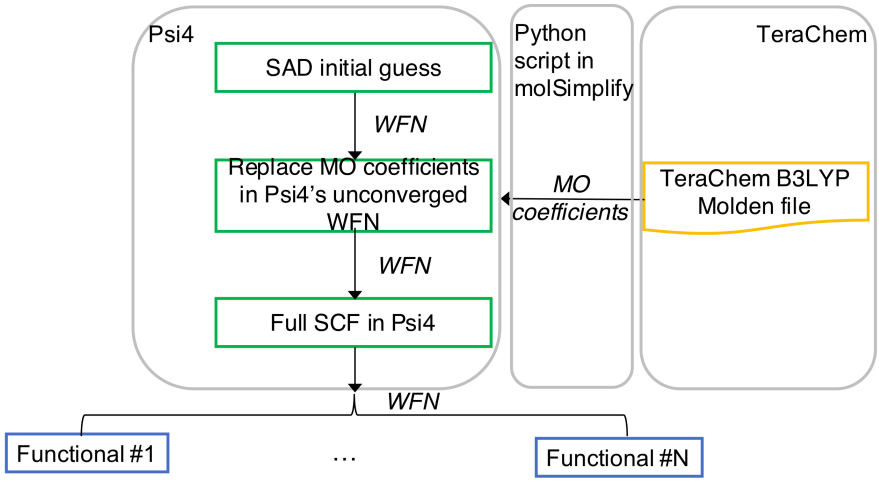

Figure S31. Computational workflow for setting up multiple single-point calculations. The molecular orbital (MO) coefficients of the B3LYP converged wavefunction obtained in TeraChem[8,9] are first extracted by a routine in our open-source package molSimplify[10,11]. These MO coefficients are used to replace the MO coefficients that would normally be obtained from an initial guess by the superposition of atomic density (SAD) for the wavefunction in Psi4[12]. A self-consistent field calculation with B3LYP is then performed to obtain the converged B3LYP wavefunction in Psi4. This wavefunction is used as the initial guess for the single-point energy calculations in the 22 functionals other than B3LYP to maximize correspondence of the electronic states converged across different functionals. For larger basis sets, a similar procedure is employed but using basis set projection inside Psi4 starting from the Psi4 B3LYP/LACVP* converged wavefunction.

Table S17. Summary of the default DFT calculation parameters and those used in this work in Psi4[13,14]. We used a smaller maximum SCF iteration because we read in the converged B3LYP wavefunction as the initial guess for all functionals. We also chose 3e-5 Ha as the density convergence threshold to be consistent with the TeraChem default setup, which was used to obtain the equilibrium geometries in *MD1* and *OHLDB*.

	This work	Default
Number of radial points	99	75
Number of spherical points	599	302
DFT pruning scheme	robust	robust
Maximum SCF iterations	50	100
Density convergence threshold	3e-5 Ha	1e-6 Ha

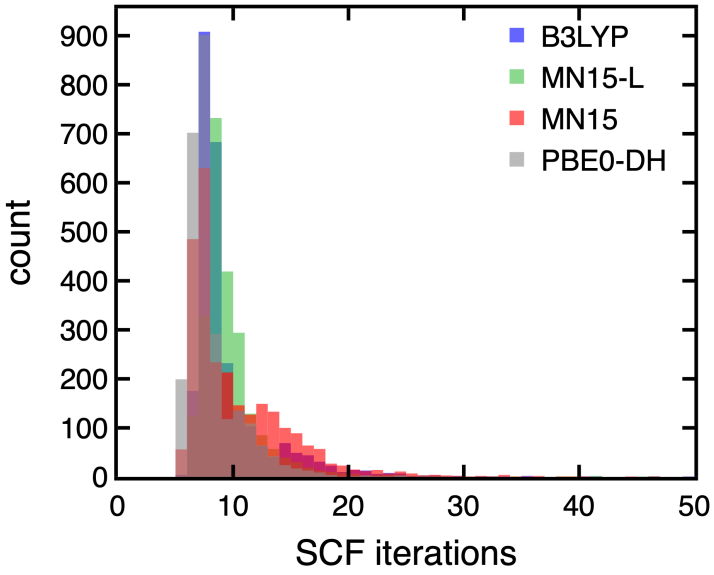

Figure S32. Unnormalized histograms of the number of SCF iterations needed for convergence of the N-electron systems calculated with the LACVP* basis set for four representative functionals, B3LYP (blue), MN15-L (green), MN15 (red), and PBE0-DH (gray). Results were obtained on the unique complexes in *MD1+OHLDB*. The bin size is one SCF iteration for all four DFAs.

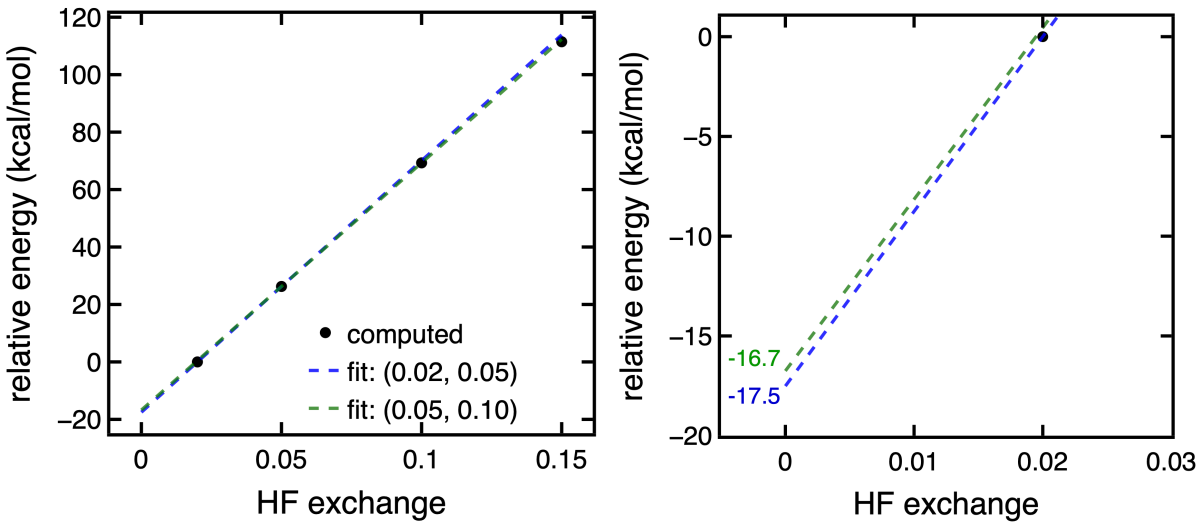

Figure S33. Example of the Hartree–Fock (HF) linear extrapolation scheme (left) and expansion highlighting the 0.00 HF exchange fraction (right) for a representative quartet $Cr(III)(CO)_4(furan)_2$ complex to obtain the TPSS/LACVP* result that did not initially converge. The extrapolated value at 0.00 HF exchange from the linear fit of TPSS with 0.02 HF exchange and TPSS with 0.05 HF exchange fraction (blue dashed line) is used as an approximation for the TPSS energy of this complex. This result would deviate slightly from the green line obtained between the 0.05 and 0.10 HF exchange fraction. The relative energy at 0.02 HF exchange fraction is set to zero for comparison.

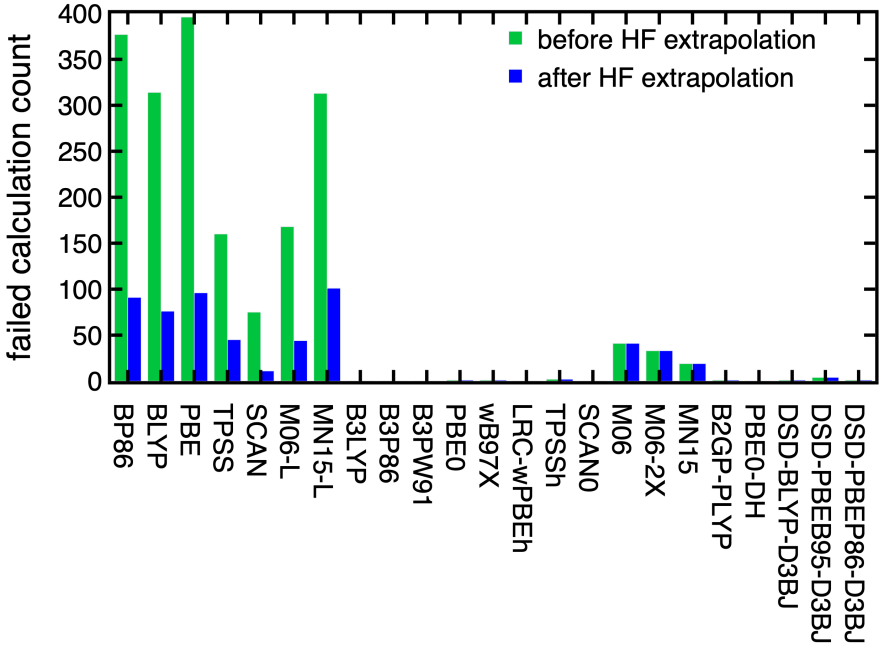

Figure S34. Numbers of calculations for which the converged SCF energy cannot be obtained with the LACVP* basis set before (green) and after (blue) HF linear extrapolation with each functional for the *N*-electron *MD1+OHLDB* complexes.

Text S2. Description of Hartree–Fock resampling procedure for converging DFT single point energies.

If a single-point calculation for a pure GGA or meta-GGA did not converge, we automatically performed single-point calculations with the hybrid version of that GGA or meta-GGA at a series of Hartree–Fock (HF) exchange percentages, starting at 15, then decreasing to 10, 5, and 2. Each subsequent HF exchange percentage calculation uses the wavefunction from the prior HF exchange percentage and is only performed if the calculation of the prior HF exchange percentage converges in the default maximum number of iterations. We then obtained the line formed by two points of the last three total energies converged (e.g., one with 10 and 5 and one with 5 and 2). We extrapolated the two lines to 0 percent HF exchange. If the two extrapolated total energy values did not deviate by more than 2.5 kcal/mol, the extrapolated value (i.e., to 0) of the last two points (i.e., 5 and 2) was used as an approximate energy for this GGA or meta-GGA. If the 2% calculation was not attempted or not converged or if the two extrapolations disagreed by > 2.5 kcal/mol, we indicated the linear extrapolation failed and removed the complex from the dataset.

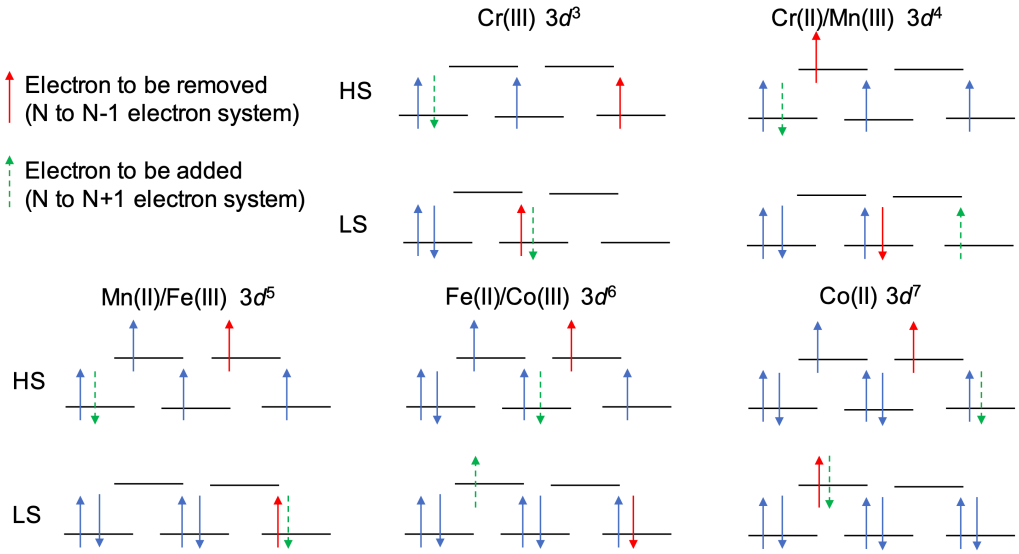

Figure S35. Conventions for adding (green dashed arrow) or removing (red solid arrow) electrons to an *N*-electron system to form the *N*+1-electron and *N*-1-electron systems. Both the high-spin (HS) and low-spin (LS) cases are shown for all *d* shell configurations (d^3 to d^7) considered in this work.

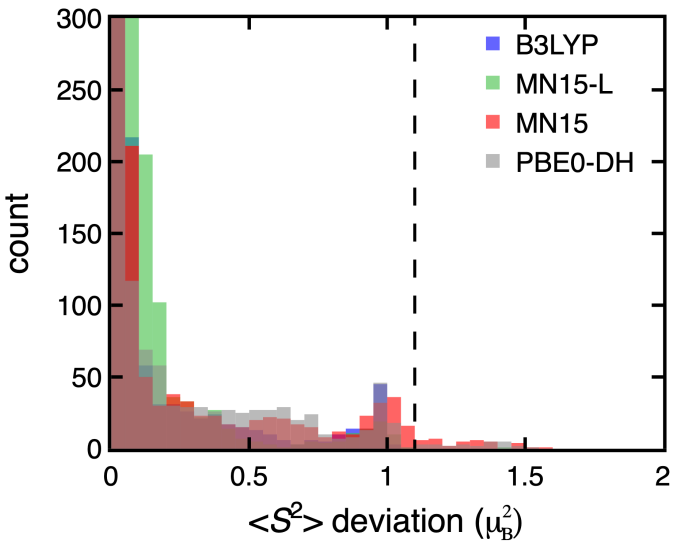

Figure S36. Histogram of $\langle S^2 \rangle$ deviations from $S(S+1)$ for the N-electron systems calculated with LACVP* basis set for four representative functionals on the unique complexes in *MD1* and *OHLDB*: B3LYP (blue), MN15-L (green), MN15 (red), and PBE0-DH (gray). The bin size is 0.05 for all 4 functionals. The cutoff value of 1.1 μ_B^2 is shown as a dashed black line.

Table S18. Number of calculations that were removed either because of the failure to obtain the final energy or an $\langle S^2 \rangle$ deviating from $S(S+1)$ by more than 1.1 μ_B^2 for the original N-electron calculation as well as the N-1 and N+1 calculations. Results are reported for both the LACVP* and def2-TZVP basis sets. In practice, most calculations were eliminated because of the failure to obtain the final self-consistent field result. Ideally, we would have 2,639 successful calculations for N-1, N, N+1 electron systems if no calculation failed (i.e., all zeros in this table).

		LACVP*			def2-TZVP		
		N-1	N	N+1	N-1	N	N+1
GGA	BP86	439	377	133	542	403	190
	BLYP	392	314	128	506	382	179
	PBE	465	397	167	551	394	194
meta-GGA	TPSS	263	170	101	363	235	125
	SCAN	161	101	79	145	103	72
	M06-L	217	169	70	154	232	66
	MN15-L	506	314	111	417	107	100
hybrid	B3LYP	98	0	34	68	2	32
	B3P86	103	8	36	72	2	41
	B3PW91	109	10	40	80	4	37
	PBE0	135	16	47	100	6	41
meta-GGA hybrid	TPSSh	126	15	37	94	7	44
	SCAN0	199	30	85	169	20	73
	M06	172	34	56	129	16	48
	M06-2X	256	25	57	154	4	57
	MN15	215	47	47	116	13	48
range-separated hybrid	LC-ωPBEh	143	17	48	101	5	50
	ωB97X	131	3	34	95	4	42
double hybrid	B2GP-PLYP	284	21	81	267	14	85
	PBE0-DH	222	27	69	200	15	70
	DSD-BLYP-D3BJ	327	27	84	292	20	95
	DSD-PBEB95-D3BJ	314	18	75	248	18	81
	DSD-PBEP86-D3BJ	343	30	86	305	18	103

Table S19. Number of data points for each property that were obtained for all 23 DFAs with each basis set combination from the unique *MD1+OHLDB* complexes.

	LACVP*	def2-TZVP	theoretical maximum
$\Delta E_{H\text{-}L}$	845	862	1068
vertical IP	1406	1447	2639
Δ-SCF gap	1214	1227	2639

Text S3. We have introduced a systematic approach to featurize molecular inorganic complexes that blends metal-centric and whole-complex topological properties in a feature set referred to as revised autocorrelation functions (RACs).[15] These RACs, variants of graph autocorrelations (ACs),[16-19] are sums of products and differences of atomic properties, i.e., electronegativity (χ), nuclear charge (Z), topology (T), covalent radius (S), and identity (I). Standard ACs are defined as

$$P_d = \sum_i \sum_j P_i P_j \delta(d_{ij}, d)$$

where P_d is the AC for property P at depth d, δ is the Dirac delta function, and d_{ij} is the bond-wise path distance between atoms i and j.

In our approach, we have five types of RACs:

- ${}^{\mathrm{f}}_{\mathrm{all}}P_d$: standard ACs start on the full molecule (*f*) and have all atoms in the scope (all).
- ${}^{\mathrm{f}}_{\mathrm{ax}}P_d$ and ${}^{\mathrm{f}}_{\mathrm{eq}}P_d$: restricted-*scope* ACs that start on the full molecule (*f*) and separately evaluate axial or equatorial ligand properties

$$ {}^{\mathrm{f}}_{\mathrm{ax/eq}}P_d = \frac{1}{\left|\text{ax/eq ligands}\right|}\sum_{i}^{n_{\mathrm{ax/eq}}}\sum_{i}^{n_{\mathrm{ax/eq}}} P_i P_j \delta(d_{ij}, d) $$

where $n_{\mathrm{ax/eq}}$ is the number of atoms in the corresponding axial or equatorial ligand and properties are averaged within the ligand subtype.

- ${}^{\mathrm{mc}}_{\mathrm{all}}P_d$: restricted-scope, metal-centered (mc) descriptors that start on the metal center (mc) and have all atoms in the scope (all), in which one of the atoms, *i,* in the *i,j* pair is a metal center:

$$ {}^{\mathrm{mc}}_{\mathrm{all}}P_d = \sum_{i}^{\mathrm{mc}}\sum_{i}^{\mathrm{all}} P_i P_j \delta(d_{ij}, d) $$

- ${}^{\mathrm{lc}}_{\mathrm{ax}}P_d$:and ${}^{\mathrm{lc}}_{\mathrm{ax}}P_d$: restricted-scope, metal-proximal ACs that start on a ligand-centered (lc) and separately evaluate axial or equatorial ligand properties, in which one of the atoms, *i*, in the *i,j* pair is the metal-coordinating atom of the ligand:

$$ {}^{\mathrm{lc}}_{\mathrm{ax/eq}}P_d = \frac{1}{\left|\text{ax/eq ligands}\right|}\frac{1}{\left|\mathrm{lc}\right|}\sum_{i}^{\mathrm{lc}}\sum_{i}^{n_{\mathrm{ax/eq}}} P_i P_j \delta(d_{ij}, d) $$

We also modify the AC definition, *P′*, to property differences rather than products for a minimum depth, d, of 1 (as the depth-0 differences are always zero):

$$ {}^{\mathrm{lc/mc}}_{\mathrm{ax/eq/all}}P'_d = \sum_{i}^{\text{lc or mc scope}}\sum_{i} P_i P_j \delta(d_{ij}, d) $$

where scope can be axial, equatorial, or all ligands.

We demonstrated these RACs to be predictive for inorganic chemistry properties, such as spin-state splitting and ionization/redox potential. Over all possible origins (i.e. metal-centered, mc, or ligand-centered, lc) there are 42*d*+30 theoretical RAC features, where *d* is the maximum distance in bond paths through which two atoms are correlated in a single descriptor.[15] After eliminating the identity product RACs, There are 30*d*+25 product-based RACs (i.e., 6*d*+6 for each property) that arise from differing starting points (e.g., metal-centered or ligand-centered). After eliminating the identity difference RACs, there are 12*d* additional nontrivial difference RACs. With a bond depth cutoff of 3, this gives 151 RACs in total. Note that a given depth cutoff does not mean that whole-molecule information is excluded since the information can be included through the summation in RACs, but it does allow the user to choose not to directly correlate in a single feature the product of properties of two atoms farther apart than a certain topological distance. In this work, the full definition of the RAC representation also included oxidation state, spin state, and total ligand charge, for a total of 154 features.

Table S20. Range of hyperparameters sampled for ANN models trained from scratch with Hyperopt[20]. The lists in the architecture row can refer to one, two, or three hidden layers (i.e., the number of items in the list), and the number of nodes in each layer are denoted as elements of the list. The built-in Tree of Parzen Estimator algorithm in Hyperopt was used for the hyperparameter selection process.

Architecture	{[128], [256], [512], [128, 128], [256, 256], [512, 512], [128, 128, 128], [256, 256, 256], [512, 512, 512]}
L2 regularization	[1e-6, 1]
Dropout rate	[0, 0.5]
Learning rate	[1e-6, 1e-3]
Beta1	[0.75, 0.99]
Batch size	[16, 32, 64, 128, 256, 512]

Table S21. Range of two hyperparameters for the radial-basis function (RBF) kernel sampled for KRR models with Hyperopt, where the regularization coefficients is the L2 regularization weight and decay width is the standard deviation of the Gaussian distribution in the RBF kernel. The built-in Tree of Parzen Estimator algorithm in Hyperopt was used for the hyperparameter selection process.

Regularization coefficient	[1e-8, 100]
Decay width	[1e-8, 100]

References

(1) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure-Property Relationships. *J. Phys. Chem. A* **2017,** *121* (46), 8939.

(2) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018,** *57* (42), 13973.

(3) Janet, J. P.; Duan, C. R.; Yang, T. H.; Nandy, A.; Kulik, H. J. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chem. Sci.* **2019,** *10* (34), 7913.

(4) Liu, F.; Duan, C. R.; Kulik, H. J. Rapid Detection of Strong Correlation with Machine Learning for Transition-Metal Complex High-Throughput Screening. *J. Phys. Chem. Lett.* **2020,** *11* (19), 8067.

(5) Taylor, M. G.; Yang, T.; Lin, S.; Nandy, A.; Janet, J. P.; Duan, C. R.; Kulik, H. J. Seeing Is Believing: Experimental Spin States from Machine Learning Model Structure Predictions. *J. Phys. Chem. A* **2020,** *124* (16), 3286.

(6) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016,** *72*, 171.

(7) Rose, M. E.; Kitchin, J. R. pybliometrics: Scriptable bibliometrics using a Python interface to Scopus. *Softwarex* **2019,** *10*.

(8) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **2009,** *5* (10), 2619.

(9) Petachem, L. PetaChem. http://www.petachem.com. (Accessed June 24, 2021).

(10) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. molSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016,** *37*, 2106.
(11) KulikGroup. molSimplify documentation. 2020. http://molsimplify.mit.edu. (Accessed June 24, 2021).
(12) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.et al. PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017,** *13* (7), 3185.
(13) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985,** *82* (1), 270.
(14) Psi4. Psi4 manual. https://psicode.org/psi4manual/master/dft.html. (Accessed June 24, 2021).
(15) Janet, J. P.; Kulik, H. J. Resolving transition metal chemical space: feature selection for machine learning and structure-property relationships. *J. Phys. Chem. A* **2017,** *121* (46), 8939.
(16) Devillers, J.; Domine, D.; Guillon, C.; Bintein, S.; Karcher, W. Prediction of partition coefficients (log p oct) using autocorrelation descriptors. *SAR QSAR Environ. Res.* **1997,** *7* (1-4), 151.
(17) Broto, P.; Devillers, J. *Autocorrelation of properties distributed on molecular graphs*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990.
(18) Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *J. Am. Chem. Soc.* **2013,** *135* (19), 7296.
(19) Broto, P.; Moreau, G.; Vandycke, C. Molecular Structures: Perception, Autocorrelation Descriptor and Sar Studies: System of Atomic Contributions for the Calculation of the N-Octanol/Water Partition Coefficients. *Eur. J. Med. Chem.* **1984,** *19* (1), 71.
(20) Bergstra, J.; Yamins, D.; Cox, D. D. Proceedings of the 12th Python in science conference, 2013; p 13.