Supporting Information for

Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design

Jon Paul Janet[1,#], Terry Z.H. Gani[1,#], Adam H. Steeves[1], Efthymios I. Ioannidis[1], and Heather J. Kulik[1,*]

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

email:hjkulik@mit.edu

Contents

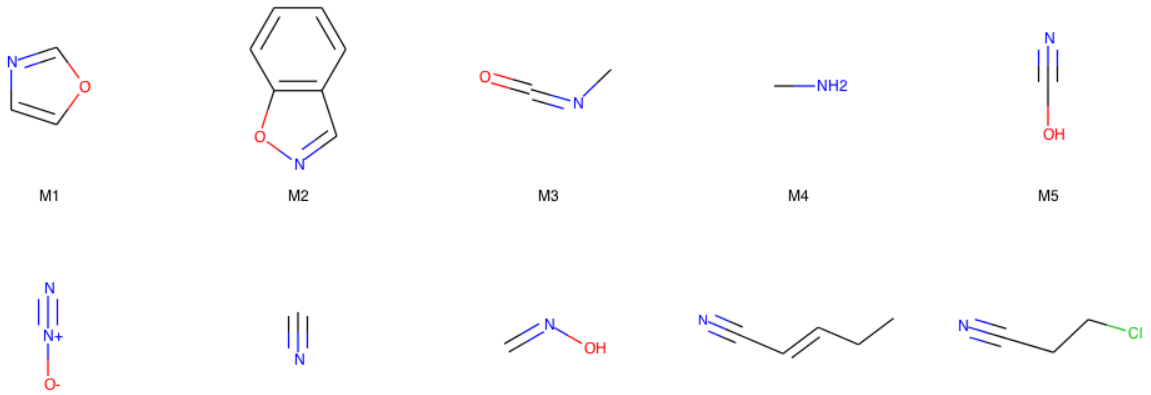

Figure S1. Structures of all 10 monodentate ligands studied in this work, with representative structures highlighted in the main text labeled “M1” through “M5”.

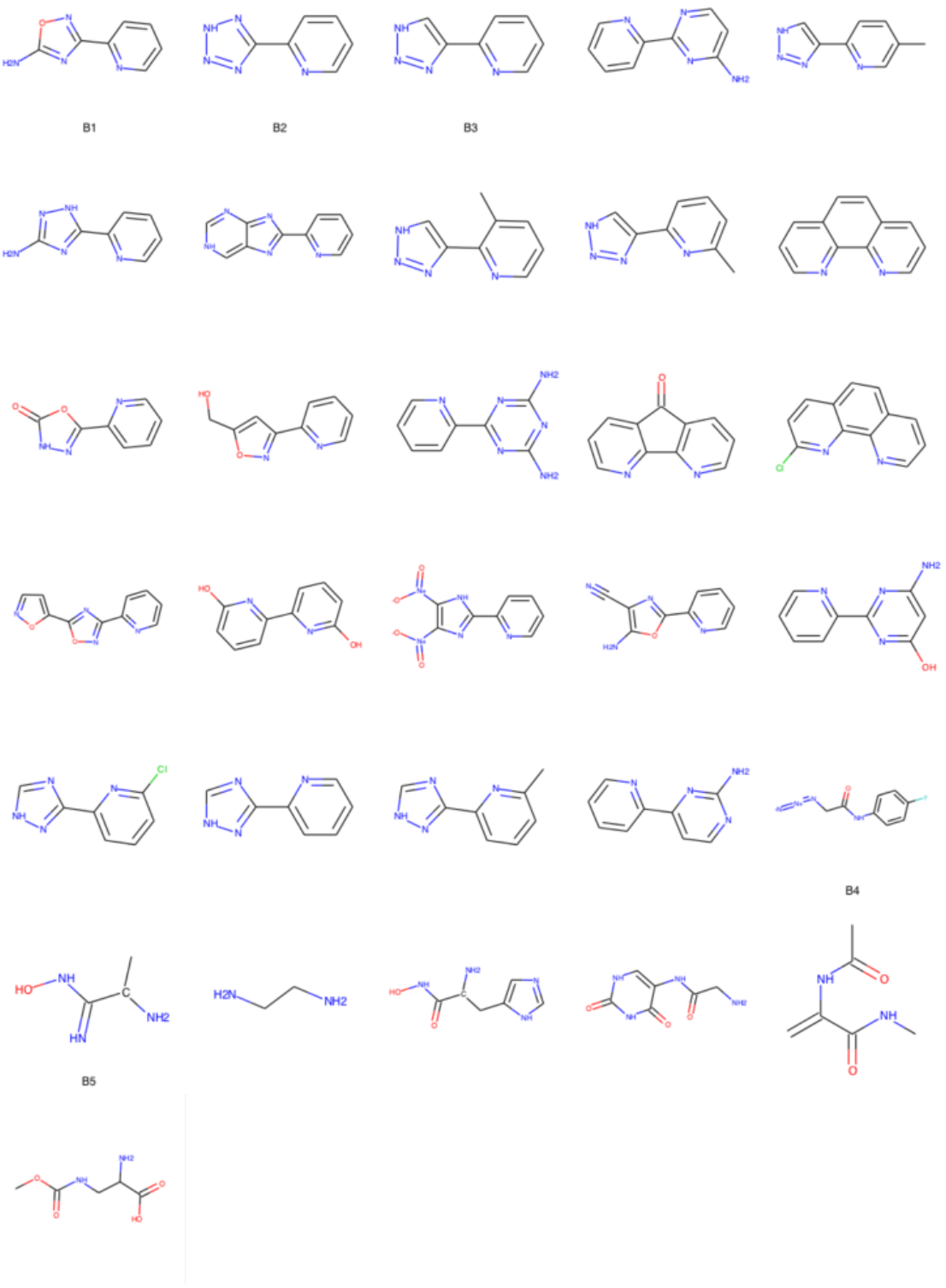

Figure S2. Structures of all 31 bidentate ligands studied in this work, with representative structures highlighted in the main text labeled “B1” through “B5”.

Table S1. Basis set dependence and dispersion effect on DFT computed ionization energies (ΔE, 2+ to 3+) of Fe complexes with representative ligands. The baseline DFT values correspond to 6-31g* and the B3LYP functional. LANL2DZ effective core potential is used for Fe in all cases. A '-' indicates convergence issues were encountered with diffuse basis functions.

Lig	Spin	ΔE (DFT), kcal/mol	ΔE(DFT+D3), kcal/mol	ΔE(6-31+g*), kcal/mol
B1	LS	280.72	279.18	284.19
	HS	287.22	285.00	290.52
B2	LS	317.04	315.70	319.36
	HS	326.96	324.72	328.73
B3	LS	281.47	280.15	284.23
	HS	291.82	289.35	294.18
B25	LS	67.39	63.03	77.14
	HS	86.86	83.34	96.92
B26	LS	35.80	34.97	44.56
	HS	53.88	54.21	61.44
M1	LS	306.01	303.60	309.24
	HS	316.91	313.15	319.39
M2	LS	269.03	267.34	-
	HS	277.94	274.83	-
M3	LS	346.07	343.14	348.11
	HS	355.62	354.03	356.93
M4	LS	317.89	317.03	318.27
	HS	333.94	332.97	333.06
M5	LS	323.58	322.69	327.05
	HS	325.35	323.74	327.29

Table S2. Summary of number of results at each step of filtering from search of the ChEMBL-21 for monodentate query [#7D1,#7D2;!+], bidentate query 1 [#7D1,#7D2;!+;a;H0]1c(c[#7D1,#7D2;!+;a;H0])cccc1, and bidentate query 2 [#7D1,#7D2;!+][#6;!R][#6;!R][#7D1,#7D2;!+].

	SMARTS query		
# matches after…	Monodentate	First bidentate (aromatic)	Second bidentate (aliphatic)
Substructure search	1282678	6190	83532
Size and single-match filters	563	59	130
Removing counterions, stereoisomers and duplicates	527	58	100
Element filter	355	39	76
Complex structure quality filter	311	39	41
Dissimilarity search	10	24	7

Table S3. Bond length prediction by ANN, database, and UFF compared to mean measured bond lengths at DFT relaxed geometries in Å.

	Ox	**Spin**	**Mean DFT Distance**	**Database**	**ANN**	**Mean UFF distance**
B1	2	1	2.02	2.10	2.03	2.09
	2	5	2.22	2.10	2.15	2.09
	3	2	1.99	2.10	2.03	2.09
	3	6	2.14	2.10	2.14	2.09
B2	2	1	2.04	2.10	2.03	1.99
	2	5	2.24	2.10	2.15	1.99
	3	2	2.01	2.10	2.03	1.99
	3	6	2.16	2.10	2.14	1.99
B3	2	1	2.02	2.10	2.03	2.04
	2	5	2.22	2.10	2.15	2.04
	3	2	1.99	2.10	2.03	2.04
	3	6	2.14	2.10	2.14	2.04
B25	2	1	2.08	2.10	2.02	2.07
	2	5	2.24	2.10	2.29	2.07
	3	2	2.01	2.10	2.02	2.07
	3	6	2.16	2.10	2.28	2.07
B26	2	1	2.04	2.10	2.03	2.04
	2	5	2.23	2.10	2.30	2.04
	3	2	1.99	2.10	2.03	2.04
	3	6	2.15	2.10	2.29	2.04
M1	2	1	2.08	2.10	2.02	1.97
	2	5	2.26	2.10	2.30	1.97

	3	2	2.03	2.10	2.03	1.97
	3	6	2.17	2.10	2.29	1.97
M2	2	1	2.02	2.10	2.02	2.13
	2	5	2.22	2.10	2.30	2.13
	3	2	1.99	2.10	2.03	2.13
	3	6	2.14	2.10	2.29	2.13
M3	2	1	2.17	2.10	2.02	2.02
	2	5	2.29	2.10	2.29	2.02
	3	2	2.07	2.10	2.02	2.02
	3	6	2.19	2.10	2.29	2.02
M4	2	1	2.12	2.10	2.03	2.05
	2	5	2.31	2.10	2.30	2.05
	3	2	2.08	2.10	2.03	2.05
	3	6	2.24	2.10	2.29	2.05
M5	2	1	1.99	2.10	2.02	2.03
	2	5	2.19	2.10	2.30	2.03
	3	2	1.94	2.10	2.03	2.03
	3	6	2.08	2.10	2.29	2.03

Table S4. Predicted spin splitting HFX sensitivity from ANN, DFT results at different HFX values and distance to ANN training data.

Lig	**Ox**	**ΔE_{H-L} (kcal/mol)**		**ANN Slope (kcal/mol·HFX)**	**Dist. To Train.**	**Interp. Error (kcal/mol·HFX)**
		20% HFX	**0% HFX**			
B1	2	-4.5	17.2	-90.9	0.33	-3.6
	3	1.9	16.8	-85.2	0.33	2.2
B2	2	-6.1	15.1	-90.7	0.34	-3.1
	3	3.8	17.2	-85.0	0.34	3.6
B3	2	2.2	24.7	-90.7	0.34	-4.3
	3	12.6	27.9	-85.0	0.34	1.7
B4	2	-19.8	-5.5	-91.0	1.00	3.8
	3	-0.4	9.2	-86.1	1.00	7.7
B5	2	-5.4	9.9	-89.0	0.58	2.5
	3	12.7	27.5	-83.3	0.58	1.8
M1	2	-3.1	18.1	-89.7	0.56	-3.3
	3	5.8	17.8	-83.3	0.56	4.6
M2	2	-22.2	-10.3	-90.5	0.76	6.2
	3	-12.6	-4.5	-84.2	0.76	8.8
M3	2	-9.2	9.1	-89.3	0.46	-0.4
	3	1.7	13.8	-83.0	0.46	4.4
M4	2	-9.0	5.7	-86.4	0.47	2.5
	3	7.0	22.0	-80.6	0.47	1.1
M5	2	-10.7	8.4	-89.2	0.76	-1.2
	3	-8.9	4.9	-82.8	0.76	2.7

Table S5. Predicted spin splitting HFX sensitivity from ANN and DFT results taken from linear regression on HFX = 0 to 0.3 for some homoleptic Fe complexes. These are drawn from the test set data, meaning the ANN was not trained on any of these examples.

Ox	Ligand	ANN Slope (kcal/mol·HFX	DFT Slope (kcal/mol·HFX)
2	H_2O	-67	-59
	oxalate	-78	-73
	NCS	-97	-86
3	H_2O	-42	-53
	NCS	-74	-62
	Cl	-72	-70
	bipy	-84	-75
	NH_3	-57	-76
		-118	-112

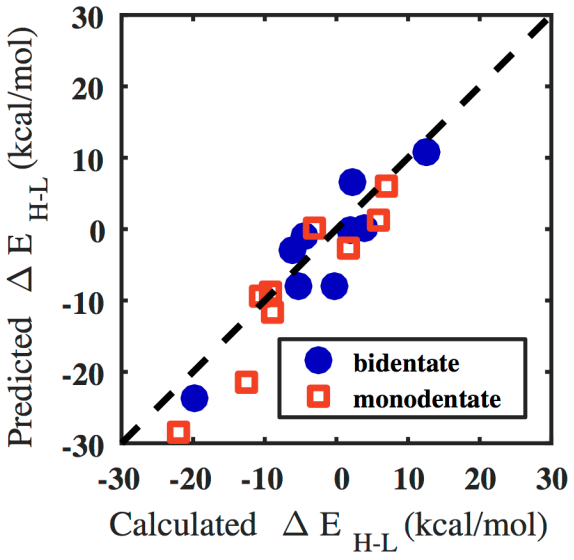

Figure S3. Comparison of spin-state splitting in kcal/mol with varying HFX exchange, predicted by ANN slope interpolation on 0-20% HFX compared to DFT results.

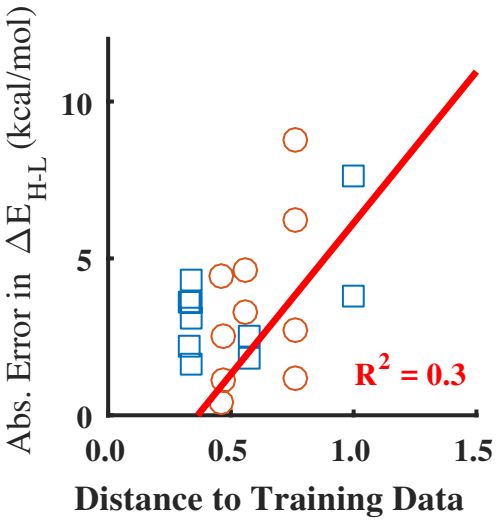

Figure S4. Correlation of absolute error in spin-state splitting interpolation in kcal/mol by the ANN slopes compared to Euclidean distance to training data in the ANN descriptor space.

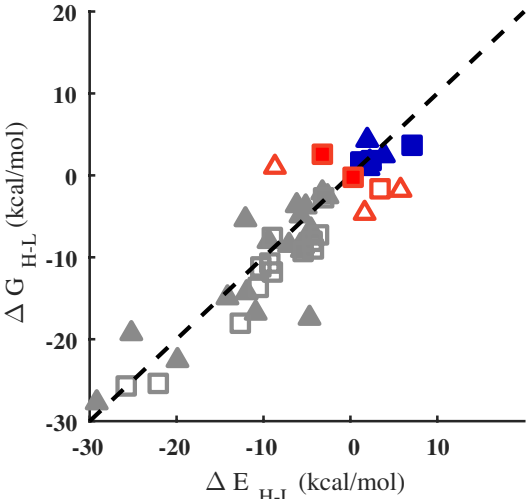

Figure S5. High-spin/low-spin spin state splitting in kcal/mol from adiabatic electronic energy differences in the gas phase compared to results after incorporation of zero point energy, entropic contribution at 300 K, and implicit solvent effects modeling water. The cases where the spin ordering changes are highlighted in red, whereas low-spin cases are in blue and high-spin in gray. Bidentate structures are marked with triangles and monodentate structures are marked with squares. Filled symbols correspond to Fe(II) complexes.

Table S6. Gas phase spin splitting energy, solvation energy, total thermochemistry corrections (including ZPE contribution) at 300 K, and spin splitting after solvation and entropic effects, shown at both high (H) and low spin (L) states and II and III oxidation. All units are kcal/mol.

Lig	Ox	ΔE_{H-L}	$\Delta G_{s,L}$	$\Delta G_{s,H}$	$\Delta E_{vib,L}$	$E_{vib,H}$	ΔG_{H-L}
1	2	-4.5	-135.5	-132.8	241.4	236.7	-6.6
1	3	1.9	-286.4	-282.3	243.5	237.7	0.3
2	2	-6.1	-173.7	-168.9	212.3	209.9	-3.7
2	3	3.8	-350.4	-346.4	213.5	209.0	3.2
3	2	2.2	-139.9	-138.2	236.8	234.4	1.6
3	3	12.6	-294.1	-290.4	238.5	233.0	10.7
4	2	-4.8	-138.1	-147.8	300.9	297.9	-17.4
4	3	6.6	-294.0	-292.9	300.4	295.2	2.5
5	2	2.1	-136.6	-135.0	285.8	282.9	0.8
5	3	12.7	-287.3	-283.8	287.5	281.0	9.6
6	2	111.1	-153.2	-151.0	266.7	263.0	109.5
6	3	3.4	-312.2	-308.5	270.0	261.2	-1.7
7	2	-3.3	-162.9	-160.8	300.5	299.6	-2.0
7	3	8.0	-309.2	-308.4	301.1	295.2	2.9
8	2	3.9	-136.6	-134.9	288.1	284.9	2.4
8	3	13.6	-287.4	-283.6	289.8	283.1	10.8
9	2	-7.1	-137.5	-135.3	288.4	284.8	-8.5
9	3	3.5	-288.9	-285.5	291.1	284.5	0.3
10	2	1.9	-122.8	-119.9	314.1	313.6	4.3
10	3	12.4	-273.6	-271.2	316.4	311.0	9.5
11	2	-5.8	-176.0	-173.1	218.7	216.7	-4.9
11	3	-2.4	-348.0	-345.1	220.2	215.4	-4.2
12	2	0.4	-134.3	-133.8	288.8	287.7	-0.3
12	3	6.4	-279.5	-275.8	287.8	283.9	6.3
13	2	-5.8	-136.5	-135.1	308.1	303.8	-8.5
13	3	4.5	-284.8	-282.0	310.5	303.5	0.3
14	2	108.6	-147.4	-143.3	267.3	266.2	111.7
14	3	-1.1	-309.8	-307.3	271.6	262.3	-8.0
15	2	-9.4	-121.8	-121.3	295.0	295.9	-8.0
15	3	2.5	-268.0	-264.9	294.3	289.6	0.9
16	2	-3.3	-149.8	-140.8	276.7	273.5	2.6
16	3	1.5	-304.0	-299.1	276.0	271.7	2.1
17	2	109.0	-136.8	-134.6	306.1	302.6	107.6
17	3	-11.8	-287.7	-286.1	304.8	304.6	-10.5
18	2	102.3	-179.8	-188.5	256.3	253.2	90.5
18	3	-0.6	-342.3	-350.1	256.0	251.9	-12.4
19	2	101.6	-149.6	-151.9	256.8	251.9	94.4
19	3	-6.7	-295.8	-291.7	256.2	255.2	-3.6
20	2	-4.5	-132.6	-131.2	307.0	303.3	-6.8
20	3	5.6	-279.4	-277.2	309.5	303.3	1.7
21	2	-14.1	-145.0	-142.6	217.0	213.7	-15.0
21	3	-2.4	-304.5	-300.0	216.8	212.2	-2.5
22	2	-2.6	-145.9	-142.5	237.7	234.4	-2.6

22	3	8.3	-309.4	-305.8	238.9	234.4	7.4
23	2	-12.0	-144.3	-142.4	290.2	286.0	-14.3
23	3	-0.1	-304.3	-301.2	289.8	284.1	-2.6
24	2	-5.6	-140.3	-138.4	301.5	296.2	-9.1
24	3	6.1	-296.8	-293.5	298.9	294.9	5.4
25	2	-19.8	-56.5	-56.1	244.6	241.6	-22.6
25	3	-0.4	-35.0	-35.3	245.6	241.6	-4.7
26	2	-5.4	-67.9	-67.9	232.6	229.4	-8.6
26	3	12.7	-34.6	-37.5	233.8	226.7	2.6
27	2	-5.1	-178.6	-172.4	215.8	210.9	-3.9
27	3	9.6	-397.5	-388.2	216.5	211.7	14.1
28	2	-12.1	-80.0	-70.5	300.5	297.7	-5.4
28	3	11.0	-55.0	-50.8	301.6	297.1	10.7
29	2	-25.2	-95.6	-88.0	265.5	263.7	-19.3
29	3	6.3	-75.3	-67.3	267.6	262.6	9.3
30	2	-29.2	-525.2	-520.6	245.5	242.3	-27.8
30	3	-2.1	-306.6	-299.4	251.2	244.7	-1.4
31	2	-10.8	-70.7	-70.4	284.0	277.7	-16.8
31	3	15.8	-45.0	-48.5	286.8	279.5	5.0
1	2	-9.2	-147.4	-145.8	212.9	209.9	-10.7
1	3	1.7	-320.8	-319.1	218.7	210.7	-4.6
2	2	-3.1	-118.6	-114.3	379.3	375.4	-2.8
2	3	5.8	-253.8	-248.8	384.7	372.0	-1.9
3	2	-22.2	-164.5	-164.0	182.6	178.9	-25.4
3	3	-12.6	-350.0	-348.7	184.7	177.9	-18.1
4	2	-9.0	-175.2	-172.8	248.9	243.7	-11.8
4	3	7.0	-385.0	-380.4	251.5	243.5	3.6
5	2	-10.7	-161.6	-159.7	74.1	69.2	-13.7
5	3	-8.9	-331.8	-325.5	76.6	71.7	-7.5
6	2	-25.8	-187.4	-183.5	32.5	28.7	-25.6
6	3	-33.8	-398.2	-388.5	39.4	33.7	-29.8
7	2	-3.7	-155.1	-151.0	64.1	56.4	-7.3
7	3	1.2	-346.9	-339.5	66.7	59.7	1.6
8	2	-10.3	-174.0	-169.8	169.2	164.1	-11.2
8	3	2.3	-352.8	-347.5	170.7	164.9	1.8
9	2	-4.4	-111.2	-109.5	380.9	375.5	-8.1
9	3	-4.3	-249.5	-243.2	385.8	374.8	-9.1
10	2	-5.5	-166.5	-161.7	227.1	218.4	-9.4
10	3	-8.7	-342.2	-331.1	225.9	224.4	1.0

Table S7. Gas phase energy, solvation energy, total thermochemistry corrections (including ZPE contribution) at 300 K, and overall redox potential values for ferrocenium redox couple reference.

	Fc	**Fc^+**
E_g, Ha	-510.5205	-510.2912
ΔG_{sol}, kcal/mol	-4.0	-48.0
ΔE_{vib}, kcal/mol	104.0	104.2
ΔG_{solv}, kcal/mol	-100.0	
E^0(RC), V	4.34	

Table S8. Redox energies in the gas phase, net solvent corrections, solvated redox energies, and redox potentials relative to ferrocenium couple.

Lig	**Ground State**	**ΔG_g, kcal/mol**	**$\Delta(\Delta G_s)$, kcal/mol**	**ΔG_{solv}, kcal/mol**	**$E - E^0$, V**
bidentate					
1	H	-288	149	-139	1.68
2	H	-326	178	-149	2.10
3	L	-283	154	-129	1.25
4	H	-291	145	-145	1.97
5	L	-277	151	-127	1.15
6	L	-287	159	-128	1.19
7	H	-269	148	-121	0.93
8	L	-278	151	-127	1.15
9	H	-291	150	-141	1.76
10	L	-273	151	-123	0.97
11	H	-316	172	-144	1.92
12	H	-280	142	-138	1.66
13	H	-282	147	-135	1.51
14	L	-300	162	-137	1.61
15	H	-282	144	-139	1.68
16	L	-291	154	-137	1.59
17	L	-301	151	-151	2.19
18	L	-303	163	-141	1.77
19	L	-282	146	-136	1.54
20	H	-286	146	-140	1.72
21	H	-305	157	-147	2.04
22	H	-298	163	-135	1.52
23	H	-297	159	-138	1.64
24	H	-292	155	-137	1.61
25	H	-87	-21	-108	0.33
26	H	-51	-30	-82	-0.80
27	H	-338	216	-122	0.96

28	H	-68	-20	-88	-0.54
29	H	-74	-21	-95	-0.22
30	H	145	-221	-76	-1.03
31	H	-70	-22	-91	-0.37
monodentate					
1	H	-318	173	-144	1.92
2	H	-275	134	-140	1.73
3	H	-355	185	-170	3.03
4	H	-334	208	-126	1.13
5	H	-328	166	-162	2.69
6	H	-408	205	-203	4.47
7	H	-372	189	-184	3.62
8	H	-339	178	-161	2.63
9	H	-280	134	-146	1.99
10	H	-322	169	-153	2.28

Table S9. Low- to low-spin and high- to high-spin redox potentials relative to reference for Fe(II) complexes showing near degenerate (<5 kcal/mol) spin splitting energy after solvent and thermos corrections.

Lig	$\Delta G_{H\text{-}L}$, kcal/mol	L-L Redox, V	H-H Redox, V
	bidentate		
2	-3.74	1.80	2.10
3	1.57	1.25	1.65
5	0.82	1.15	1.53
7	-2.05	0.71	0.93
8	2.41	1.15	1.52
10	4.26	0.97	1.20
11	-4.87	1.89	1.92
12	-0.26	1.38	1.66
16	2.56	1.59	1.57
22	-2.60	1.08	1.52
27	-3.86	0.18	0.96
	monodentate		
2	-2.81	1.69	1.73

Table S10. Normalized descriptor values for all complexes, where d is denticity, mΔχ is the maximum electronegativity difference, K and tK are the Kier and truncated Kier indices respectively, and the reaming terms are autocorrelation functions: χ_d is electronegativity, Z_d is nuclear charge, I_d is identity and T_d is topology.

lig	d	mΔχ	K	tK	χ						Z					
					0	1	2	3	4	5	0	1	2	3	4	5
	bidentate															
1	0.6	-0.26	-1.21	-1.08	0.31	0.52	0.41	0.33	0.21	0.24	0.31	0.50	0.40	0.31	0.25	0.18
2	0.6	-0.26	-0.10	0.01	-0.04	0.08	0.01	-0.25	-0.25	-0.19	0.05	0.08	0.06	-0.18	-0.19	-0.20
3	0.6	-0.26	-0.10	0.01	0.02	0.14	0.10	-0.01	-0.15	-0.03	0.11	0.17	0.17	0.00	-0.07	-0.04
4	0.6	-0.26	0.54	0.64	0.59	0.65	0.60	0.68	0.54	0.65	0.66	0.69	0.63	0.69	0.59	0.71
5	0.6	-0.26	0.05	0.15	0.42	0.53	0.55	0.40	0.63	0.40	0.44	0.51	0.52	0.45	0.49	0.31
6	0.6	-0.26	0.05	0.15	0.37	0.47	0.42	0.29	0.24	0.31	0.38	0.42	0.41	0.26	0.25	0.28
7	0.6	-0.26	0.54	0.78	0.86	1.02	1.07	0.93	0.70	0.55	1.04	1.11	1.15	1.01	0.84	0.78
8	0.6	-0.26	0.05	0.15	0.42	0.53	0.55	0.40	0.67	0.92	0.44	0.51	0.52	0.47	0.69	0.82
9	0.6	-0.26	0.05	0.15	0.42	0.53	0.55	0.42	0.43	0.65	0.44	0.51	0.52	0.42	0.48	0.61
10	0.6	-0.26	0.10	0.20	0.61	0.81	0.93	1.18	1.17	1.08	0.91	1.05	1.18	1.44	1.38	1.14
11	0.6	-0.26	0.05	0.15	0.26	0.25	0.30	0.07	-0.15	0.05	0.25	0.25	0.26	0.06	-0.08	0.05
12	0.6	-0.26	0.54	0.15	0.72	0.73	0.73	0.71	0.71	0.48	0.64	0.68	0.68	0.60	0.62	0.51
13	0.6	-0.26	0.69	0.64	0.89	0.93	0.81	0.78	0.72	1.23	0.86	0.84	0.76	0.81	0.86	1.21
14	0.6	-0.26	-0.14	-0.21	0.50	0.61	0.78	0.93	0.88	0.52	0.78	0.88	1.06	1.20	1.08	0.61
15	0.6	-0.26	0.27	0.37	0.74	0.86	1.00	1.26	1.25	1.18	1.22	1.19	1.35	1.64	1.60	1.46
16	0.6	-0.26	1.00	0.15	1.10	1.17	1.10	0.78	0.78	0.79	1.18	1.18	1.13	0.77	0.90	0.96
17	0.6	-0.26	0.69	0.78	0.88	0.85	0.74	1.00	0.88	1.05	0.86	0.85	0.82	0.98	0.91	1.08
18	0.6	-0.26	1.47	0.93	1.48	1.20	1.15	1.19	1.56	1.50	1.29	1.08	1.12	1.18	1.52	1.46
19	0.6	-0.26	0.69	0.78	0.64	0.62	0.55	0.54	0.54	0.43	0.76	0.68	0.63	0.61	0.63	0.58
20	0.6	-0.26	0.69	0.64	0.89	0.89	0.77	0.91	0.85	1.07	0.86	0.85	0.79	0.91	0.88	1.08
21	0.6	-0.26	0.05	0.15	0.15	0.19	0.21	0.01	-0.14	0.03	0.43	0.31	0.34	0.15	0.10	0.19
22	0.6	-0.26	-0.10	0.01	0.02	0.14	0.13	-0.07	-0.22	-0.08	0.11	0.17	0.17	-0.05	-0.13	-0.07
23	0.6	-0.26	0.05	0.15	0.42	0.52	0.58	0.37	0.36	0.60	0.44	0.51	0.52	0.37	0.42	0.57
24	0.6	-0.26	0.54	0.64	0.59	0.65	0.60	0.66	0.53	0.84	0.66	0.69	0.63	0.69	0.65	0.87
25	0.6	-0.26	2.00	2.89	0.88	0.50	0.46	0.68	0.44	0.50	0.72	0.50	0.40	0.51	0.40	0.53
26	0.6	2.08	-0.49	-0.38	-0.33	-0.43	-0.54	-0.19	0.03	-0.53	-0.68	-0.70	-0.72	-0.45	-0.47	-0.93
27	0.6	2.08	-0.54	-0.42	-1.02	-0.99	-0.90	-0.76	-1.05	-1.39	-1.31	-1.20	-1.15	-1.12	-1.30	-1.45
28	0.6	2.08	0.84	0.25	0.81	0.67	0.64	0.92	1.33	1.14	0.45	0.40	0.42	0.58	0.81	0.79
29	0.6	2.08	0.94	1.45	0.87	0.59	0.65	0.86	0.47	0.47	0.54	0.40	0.41	0.57	0.36	0.42
30	0.6	-0.26	0.75	0.84	0.23	-0.08	-0.03	-0.23	0.01	0.39	-0.03	-0.18	-0.32	-0.27	-0.04	0.20
31	0.6	2.08	1.38	1.45	0.78	0.39	0.51	0.45	1.05	0.25	0.20	0.05	-0.02	0.15	0.38	0.06
	monodentate															
1	-1.8	-0.26	-1.67	-1.54	-1.40	-1.32	-1.26	-1.59	-1.69	-1.64	-1.36	-1.28	-1.25	-1.57	-1.67	-1.57
2	-1.8	-0.26	-1.03	-1.24	-0.51	-0.35	-0.30	-0.31	-0.44	-1.02	-0.37	-0.24	-0.15	-0.28	-0.55	-1.17
3	-1.8	-0.26	-0.54	-0.42	-1.57	-1.65	-1.64	-1.76	-1.54	-1.64	-1.58	-1.61	-1.66	-1.64	-1.57	-1.57
4	-1.8	2.08	-2.72	-2.56	-1.79	-1.72	-1.65	-1.72	-1.81	-1.64	-1.90	-1.78	-1.74	-1.74	-1.72	-1.57
5	-1.8	-0.26	-1.26	-1.14	-1.98	-2.02	-2.12	-1.97	-1.81	-1.64	-1.90	-1.95	-1.95	-1.84	-1.72	-1.57

6	-1.8	-3.55	-1.26	-1.14	-2.03	-2.11	-2.19	-2.04	-1.81	-1.64	-1.97	-2.04	-2.00	-1.89	-1.72	-1.57
7	-1.8	-0.26	-2.72	-2.56	-2.27	-2.25	-2.24	-2.04	-1.81	-1.64	-2.10	-2.11	-2.05	-1.89	-1.72	-1.57
8	-1.8	-0.26	-1.26	-1.14	-1.73	-1.74	-1.88	-1.82	-1.69	-1.64	-1.80	-1.79	-1.82	-1.74	-1.67	-1.57
9	-1.8	-0.26	0.92	-0.42	-0.89	-0.95	-0.88	-0.76	-0.90	-0.95	-0.91	-0.92	-0.97	-0.93	-0.99	-1.03
10	-1.8	-0.26	0.19	-0.42	-1.33	-1.43	-1.26	-1.23	-1.54	-1.64	-1.14	-1.29	-1.23	-1.24	-1.47	-1.57

	I						**T**					
	0	1	2	3	4	5	0	1	2	3	4	5
	bidentate											
1	0.24	0.44	0.30	0.28	0.17	0.19	0.63	0.87	0.59	0.47	0.47	0.28
2	-0.10	-0.01	-0.07	-0.24	-0.29	-0.25	-0.05	-0.06	-0.03	-0.17	-0.21	-0.24
3	0.07	0.14	0.11	0.06	-0.14	0.02	0.12	0.14	0.18	0.02	-0.01	0.00
4	0.75	0.74	0.67	0.73	0.63	0.72	0.69	0.65	0.62	0.72	0.65	0.80
5	0.58	0.59	0.67	0.50	0.79	0.54	0.63	0.61	0.64	0.59	0.55	0.42
6	0.41	0.44	0.39	0.28	0.25	0.37	0.40	0.38	0.39	0.32	0.35	0.36
7	0.92	1.04	1.04	0.95	0.71	0.54	1.03	1.05	1.12	1.00	0.89	0.92
8	0.58	0.59	0.67	0.50	0.79	1.07	0.63	0.61	0.64	0.68	0.91	0.96
9	0.58	0.59	0.67	0.50	0.56	0.81	0.63	0.61	0.59	0.54	0.70	0.78
10	0.92	1.04	1.13	1.40	1.40	1.34	1.08	1.18	1.41	1.70	1.62	1.32
11	0.07	0.14	0.11	-0.02	-0.21	-0.07	0.12	0.12	0.14	-0.04	-0.08	0.00
12	0.75	0.74	0.76	0.73	0.71	0.54	0.74	0.74	0.73	0.65	0.64	0.66
13	0.92	0.89	0.76	0.73	0.71	1.16	0.80	0.72	0.64	0.82	0.99	1.24
14	0.58	0.74	0.85	0.95	0.87	0.54	0.80	0.92	1.18	1.20	1.06	0.54
15	0.92	1.04	1.13	1.40	1.40	1.34	1.08	1.18	1.41	1.70	1.62	1.32
16	0.92	1.04	0.94	0.65	0.56	0.63	0.97	0.96	0.98	0.68	0.77	0.94
17	0.92	0.89	0.76	1.02	0.94	1.07	0.80	0.81	0.83	0.98	0.98	1.14
18	0.92	0.89	0.85	0.88	1.10	1.07	0.86	0.85	0.91	0.97	0.99	1.00
19	0.58	0.59	0.48	0.43	0.48	0.37	0.51	0.49	0.51	0.47	0.45	0.52
20	0.92	0.89	0.76	0.88	0.87	0.98	0.80	0.76	0.74	0.89	0.91	1.10
21	0.07	0.14	0.11	-0.02	-0.21	-0.07	0.12	0.12	0.14	-0.04	-0.08	0.00
22	0.07	0.14	0.11	-0.02	-0.21	-0.07	0.12	0.12	0.14	-0.04	-0.08	0.00
23	0.58	0.59	0.67	0.43	0.48	0.72	0.63	0.58	0.55	0.49	0.64	0.72
24	0.75	0.74	0.67	0.73	0.63	0.98	0.69	0.65	0.62	0.76	0.82	0.98
25	0.58	0.44	0.39	0.50	0.33	0.37	0.40	0.34	0.20	0.21	0.16	0.32
26	-0.27	-0.45	-0.44	-0.16	0.10	-0.43	-0.45	-0.46	-0.53	-0.47	-0.66	-1.05
27	-0.78	-0.90	-0.72	-0.54	-0.90	-1.31	-0.79	-0.77	-0.92	-1.07	-1.30	-1.45
28	0.75	0.59	0.67	0.95	1.40	1.07	0.63	0.67	0.71	0.70	0.77	0.62
29	0.58	0.44	0.48	0.65	0.25	0.28	0.46	0.43	0.26	0.19	0.13	0.38
30	0.24	-0.01	0.02	-0.39	-0.06	0.37	0.00	-0.24	-0.57	-0.44	-0.08	0.04
31	0.58	0.29	0.39	0.36	0.87	0.19	0.34	0.21	-0.03	0.03	0.02	0.04
	monodentate											
1	-1.46	-1.35	-1.37	-1.58	-1.67	-1.66	-1.36	-1.35	-1.28	-1.53	-1.61	-1.53
2	-0.44	-0.31	-0.26	-0.24	-0.37	-0.95	-0.28	-0.19	-0.03	-0.19	-0.59	-1.23

3	-1.63	-1.65	-1.64	-1.80	-1.60	-1.66	-1.65	-1.71	-1.72	-1.64	-1.59	-1.53
4	-1.63	-1.65	-1.55	-1.65	-1.83	-1.66	-1.59	-1.62	-1.69	-1.69	-1.64	-1.53
5	-2.13	-2.09	-2.20	-2.02	-1.83	-1.66	-2.16	-2.11	-1.95	-1.77	-1.64	-1.53
6	-2.30	-2.24	-2.29	-2.10	-1.83	-1.66	-2.27	-2.20	-1.99	-1.78	-1.64	-1.53
7	-2.30	-2.24	-2.29	-2.10	-1.83	-1.66	-2.27	-2.20	-1.99	-1.78	-1.64	-1.53
8	-1.79	-1.79	-1.92	-1.87	-1.67	-1.66	-1.87	-1.89	-1.81	-1.68	-1.61	-1.53
9	-0.61	-0.75	-0.63	-0.54	-0.75	-0.87	-0.68	-0.71	-0.88	-0.93	-1.05	-1.23
10	-1.29	-1.35	-1.18	-1.21	-1.60	-1.66	-1.25	-1.26	-1.41	-1.52	-1.59	-1.53

Table S11. Scaling factors for descriptors used in redox potential model along with importance rankings as determined by the RFE algorithm. The 4 top ranked descriptors are highlighted in bold.

Name	Mean	Variance	Ranking
denticity	1.76	0.18	24
max $\Delta\chi$	0.53	0.02	20
Kier index	3.74	1.89	18
trunc. Kier index	3.59	1.96	23
χ_0	110.88	1576.31	6
χ_1	233.52	8429.11	10
χ_2	346.73	22069.20	22
χ_3	369.85	32728.02	2
χ_4	313.78	30217.76	13
χ_5	248.13	22841.05	25
Z_0	6.91	7.15	9
Z_1	**16.20**	**46.93**	**1**
Z_2	22.45	113.65	15
Z_3	21.88	134.65	16
Z_4	17.98	109.08	12
Z_5	14.03	79.85	14
I_0	16.59	34.78	21
I_1	**34.10**	**180.28**	**1**
I_2	51.56	468.39	3
I_3	56.44	724.20	7
I_4	47.51	675.47	19
I_5	37.66	516.08	17
T_0	**85.90**	**1236.09**	**1**
T_1	**205.51**	**8084.05**	**1**
T_2	265.90	17498.53	4
T_3	243.66	18698.71	5
T_4	193.17	13805.85	8
T_5	152.20	9902.40	11

Text S1. Discussion of regression strategy

While autocorrelations allow us to construct feature spaces easily, they contain redundant information and may be prone to spurious correlations, which makes a feature selection method essential. We use the average squared LOOCV error to select features as it serves as measure of model robustness, being only calculated on withheld data points. The LOOCV error decreases with feature space dimension up until the benefit of adding more features in outweighed by overfitting, at which point the error increases again. Using the overall mean-squared error or correlation coefficient to score features would lead to a monotonic improvement with feature space dimension (see Figures S6 and S7).

When we supplied the redox data set (Table S6) to molSimplify, the code returned that the optimal number of features from the LOOCV procedure was four, corresponding to a mean-squared error (MSE) of 0.43 V^2 on the full data and a R^2 value of 0.6 (main text Figure 8), but with an average LOOCV error of 0.63 V^2 on withheld points. By contrast, the linear model trained on the full space shows a reduced MSE and improved R^2 at 0.91 and 0.1 V^2 respectively, but an average LOOCV error > 120 V^2 on held-out points (see Supporting Information Figure S7). This emphasizes the necessity of using a robust measure of model performance.

The proposed model is given in as:

$$y_{\mathrm{mod}} = 3.62Z_1 - 0.36I_1 - 7.25T_0 + 3.52T_1 + 1.51 \quad \text{(S1)}$$

The full list of descriptors and their relative rankings are given in the Supporting Information (Table S7). The low ranking (20^{th}) of the max $\Delta\chi$ term, which was important for fitting the ANN data, can be explained by the uniformity of the connecting atom environment, which is N-C ($\Delta\chi$

= 0.49) except for 6/41 cases. Further, the first excluded variable is χ_3, so electronegativity information was still present until the very last step. Since the variable elimination procedure will eliminate highly correlated descriptors, and it is reasonable to assume that the various topological indicators all encode similar information (as both the kier index and the T_d value increase with the level of branchedness), only a small set are expected to be retained.

To explore the trade-off between increased redox potential from the presence of heavy heteroatoms (driving Z_1 up), versus larger and more branched structures, we combined all of the topological descriptors into a single S descriptor,

$$S = \frac{-0.36I_1 - 7.25T_0 + 3.25T_1}{(-0.36 - 7.25 + 3.25)} \quad \text{(S2)}$$

and characterize all points as negative, middle or positive based on the values of the normalized descriptors Z_1 and S (with cutoff determined by the 25th and 75th percentiles each). It is apparent that all cases with DFT redox potential < 1 V are in, or on the border of, the middle zone with respect to both variables (Figure S8), while all complexes with redox potentials > 2 V are located in the negative-negative (7 cases) or positive-positive regions (1).

This suggests that the redox potential is maximized for ligands that have large Z_1 and negative S values (the descriptors are normalized to [-1, 1] so negative values are expected and correspond to relatively low values of S). However, these descriptors are not uncorrelated, as the larger the structure, the higher all autocorrelation values will be. Z_1 can only be increased at constant S by replacing atoms with elements that have higher nuclear charge. Complexes that have the highest redox potentials (>2 V) are all observed to have $S < Z$, and the difference quantity Z - S is observed to correlate approximately ($R^2 = 0.4$) with DFT redox potentials (Figure S9)

Table S12. Model and DFT redox potentials and optimal descriptor values.

	$E - E^0$, V		Normalized Descriptors					
Lig	**4D-Model**	**DFT**	Z_1	I_1	T_0	T_1	**S**	Z_1-S
bidentate								
1	1.70	1.68	0.50	0.44	0.63	0.87	0.40	0.10
2	1.99	2.10	0.08	-0.01	-0.05	-0.06	-0.04	0.13
3	1.74	1.25	0.17	0.14	0.12	0.14	0.10	0.07
4	1.06	1.97	0.69	0.74	0.69	0.65	0.72	-0.03
5	0.74	1.15	0.51	0.59	0.63	0.61	0.64	-0.13
6	1.32	1.19	0.42	0.44	0.40	0.38	0.42	0.00
7	1.43	0.93	1.11	1.04	1.03	1.05	1.01	0.11
8	0.74	1.15	0.51	0.59	0.63	0.61	0.64	-0.13
9	0.74	1.76	0.51	0.59	0.63	0.61	0.64	-0.13
10	1.27	0.97	1.05	1.04	1.08	1.18	0.99	0.06
11	1.92	1.92	0.25	0.14	0.12	0.12	0.12	0.13
12	0.91	1.66	0.68	0.74	0.74	0.74	0.74	-0.07
13	0.98	1.51	0.84	0.89	0.80	0.72	0.88	-0.03
14	1.86	1.61	0.88	0.74	0.80	0.92	0.69	0.19
15	1.77	1.68	1.19	1.04	1.08	1.18	0.99	0.20
16	1.76	1.59	1.18	1.04	0.97	0.96	0.98	0.20
17	1.33	2.19	0.85	0.89	0.80	0.81	0.80	0.05
18	1.88	1.77	1.08	0.89	0.86	0.85	0.86	0.21
19	1.78	1.54	0.68	0.59	0.51	0.49	0.54	0.14
20	1.15	1.72	0.85	0.89	0.80	0.76	0.84	0.01
21	2.16	2.04	0.31	0.14	0.12	0.12	0.12	0.19
22	1.66	1.52	0.17	0.14	0.12	0.12	0.12	0.06
23	0.66	1.64	0.51	0.59	0.63	0.58	0.66	-0.15
24	1.06	1.61	0.69	0.74	0.69	0.65	0.72	-0.03
25	1.47	0.33	0.50	0.44	0.40	0.34	0.46	0.05
26	0.80	-0.80	-0.70	-0.45	-0.45	-0.46	-0.44	-0.25
27	0.53	0.96	-1.20	-0.90	-0.79	-0.77	-0.82	-0.38
28	0.57	-0.54	0.40	0.59	0.63	0.67	0.59	-0.19
29	0.99	-0.22	0.40	0.44	0.46	0.43	0.48	-0.08
30	0.01	-1.03	-0.18	-0.01	0.00	-0.24	0.21	-0.39
31	-0.19	-0.37	0.05	0.29	0.34	0.21	0.46	-0.41
monodentate								
1	2.50	1.92	-1.28	-1.35	-1.36	-1.35	-1.37	0.09
2	2.09	1.73	-0.24	-0.31	-0.28	-0.19	-0.36	0.11
3	2.19	3.03	-1.61	-1.65	-1.65	-1.71	-1.59	-0.02
4	1.47	1.13	-1.78	-1.65	-1.59	-1.62	-1.57	-0.21
5	3.43	2.69	-1.95	-2.09	-2.16	-2.11	-2.20	0.24
6	3.67	4.47	-2.04	-2.24	-2.27	-2.20	-2.34	0.30
7	3.41	3.62	-2.11	-2.24	-2.27	-2.20	-2.34	0.22
8	2.63	2.63	-1.79	-1.79	-1.87	-1.89	-1.86	0.07
9	0.89	1.99	-0.92	-0.75	-0.68	-0.71	-0.66	-0.26
10	1.92	2.28	-1.29	-1.35	-1.25	-1.26	-1.25	-0.05

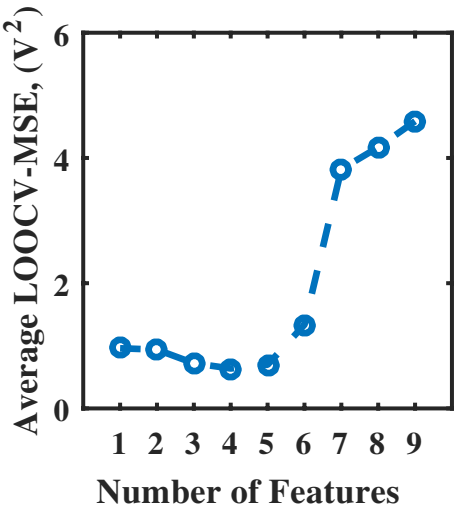

Figure S6. LOOCV-MSE with increasing number of features showing a minimum at 4 features.

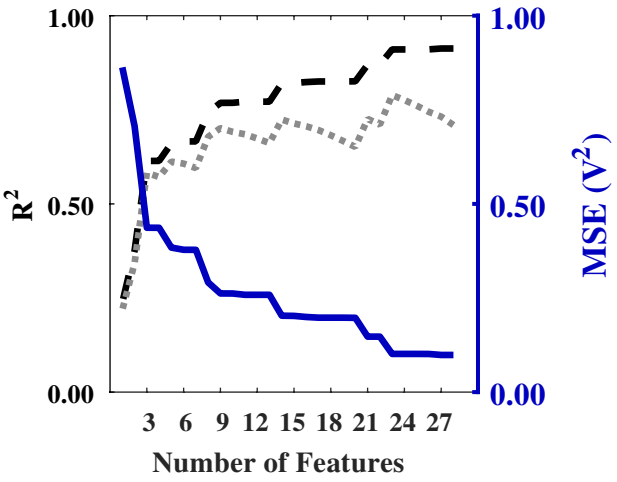

Figure S7. Correlation coefficient and MSE over whole data set with increasing number of features showing increasingly good fit with all features included. The dash line indicates the R^2 and the dotted line indicates the parameter adjusted R^2 value.

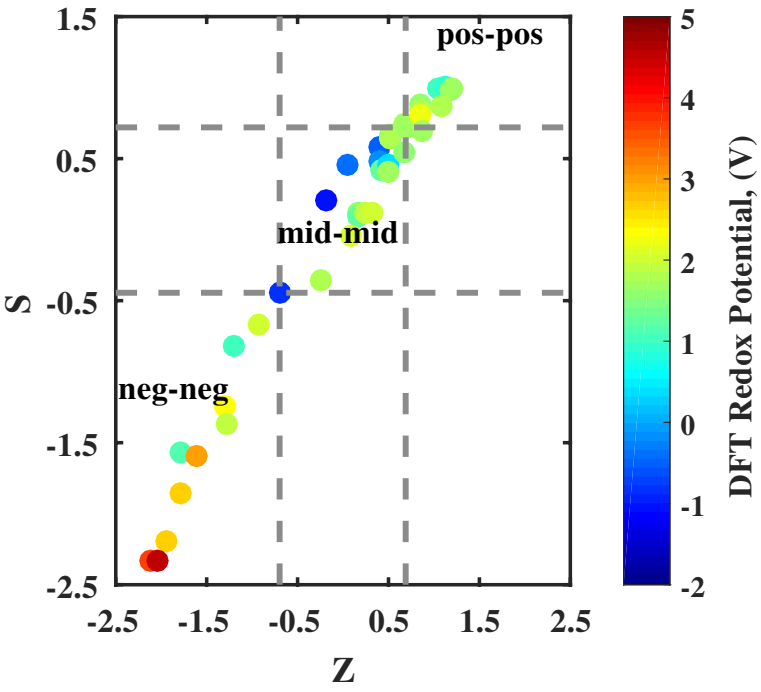

Figure S8. Plot of descriptors *S* vs. *Z* for 41 complexes colored by DFT redox potential showing that mid-range *S* and *Z* values correspond to low redox potential, whereas very negative (i.e., low) *S* and *Z* correspond to high redox potential. The signs of each range of S and Z (pos, mid, neg) are indicated on the graph.

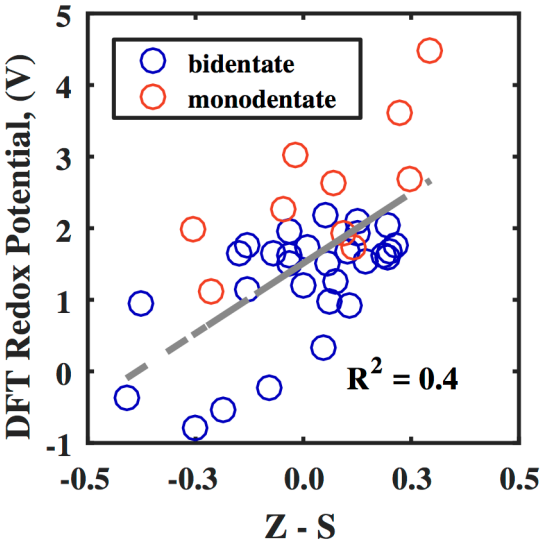

Figure S9. Plot of Z-S score vs. DFT redox potential for 41 complexes, showing correlation coefficient.