Supplementary information: Combining Hammett σ constants for Δ-machine learning and catalyst discovery

V. Diana Rakotonirina, Marco Bragato, Stefan Heinen, O. Anatole von Lilienfeld

S1 Construction of the system of linear equations

To solve for the metal constants $\boldsymbol{\rho}$, the Hammett-inspired product model (HIP) follows the blueprint laid out in Eq. 4 to build a system of linear equations $\mathbf{C}\boldsymbol{\rho} = \mathbf{0}$. Accounting for all permutations of m and n, $m \neq n$, this gives us

$$\begin{array}{c} m=1,n=2 \\ m=1,n=3 \\ \vdots \\ m=1,n=X-1 \\ m=1,n=X \\ m=2,n=1 \\ m=2,n=3 \\ \vdots \\ m=2,n=X-1 \\ m=2,n=X \\ m=3,n=1 \\ m=3,n=2 \\ \vdots \\ m=X,n=X-2 \\ m=X,n=X-1 \end{array} \begin{bmatrix} -1 & c_{12} & 0 & \dots & 0 & 0 & 0 \\ -1 & 0 & c_{13} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -1 & 0 & 0 & \dots & 0 & c_{1X-1} & 0 \\ -1 & 0 & 0 & \dots & 0 & 0 & c_{1X} \\ c_{21} & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & c_{23} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & -1 & 0 & \dots & 0 & c_{2X-1} & 0 \\ 0 & -1 & 0 & \dots & 0 & 0 & c_{2X} \\ c_{31} & 0 & -1 & \dots & 0 & 0 & 0 \\ 0 & c_{32} & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c_{XX-2} & 0 & -1 \\ 0 & 0 & 0 & \dots & 0 & c_{XX-1} & -1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_{X-2} \\ \rho_{X-1} \\ \rho_X \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{S1}$$

Constraining the first component of the vector of ρs to be 1 results in a vector $\mathbf{r}$ of length $(X-1)$ containing the remaining ρ values. HIP thus avoids trivial solutions, transforming the homogeneous system in Eq. S1 into $\mathbf{C'r} = \mathbf{c}$ (Eq. S2), where $\mathbf{c}$ is the first column of the matrix in Eq. S1 and **C'** is the remaining part. The equation to be solved for ρs then becomes

$$\begin{array}{c} m=1,n=2 \\ m=1,n=3 \\ \vdots \\ m=1,n=X-1 \\ m=1,n=X \\ m=2,n=1 \\ m=2,n=3 \\ \vdots \\ m=2,n=X-1 \\ m=2,n=X \\ m=3,n=1 \\ m=3,n=2 \\ \vdots \\ m=X,n=X-2 \\ m=X,n=X-1 \end{array} \begin{bmatrix} c_{12} & 0 & \dots & 0 & 0 & 0 \\ 0 & c_{13} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & c_{1X-1} & 0 \\ 0 & 0 & \dots & 0 & 0 & c_{1X} \\ -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & c_{23} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -1 & 0 & \dots & 0 & c_{2X-1} & 0 \\ -1 & 0 & \dots & 0 & 0 & c_{2X} \\ 0 & -1 & \dots & 0 & 0 & 0 \\ c_{32} & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & c_{XX-2} & 0 & -1 \\ 0 & 0 & \dots & 0 & c_{XX-1} & -1 \end{bmatrix} \begin{bmatrix} \rho_2 \\ \rho_3 \\ \vdots \\ \rho_{X-2} \\ \rho_{X-1} \\ \rho_X \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ -c_{21} \\ 0 \\ \vdots \\ 0 \\ 0 \\ -c_{31} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{S2}$$

S2 Other combination rules

The other combination rules that were compared to the additive one are geometric averaging (Eq. S3) [1], a sixth-power average proposed by Waldman and Hagler (Eq. S4) [2], and a harmonic mean proposed by Fender and Halsey (Eq. S5) [3].

$$\sigma_l = \epsilon \sqrt{\prod_{i=1}^{N_i} \sigma i} \tag{S3}$$

$$\sigma_l = \epsilon \left(\frac{\sum_{i=1}^{N_i} \sigma_i^6}{N_i} \right)^{1/6} \tag{S4}$$

$$\sigma_l = \epsilon \frac{N_i \prod_{i=1}^{N_i} \sigma i}{\sum_{i=1}^{N_i} \sigma_i} \tag{S5}$$

Given that the relative binding energies can be negative, another parameter ϵ was added, which had the value 1 if σ_l was positive, and -1 otherwise. Results of their test are shown in the next section.

S3 Supplementary figures

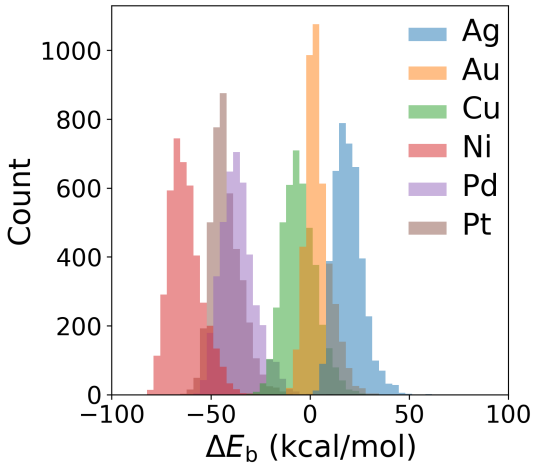


Fig. S1: Distribution of relative binding energies by metal for all 25k catalysts in DB1.

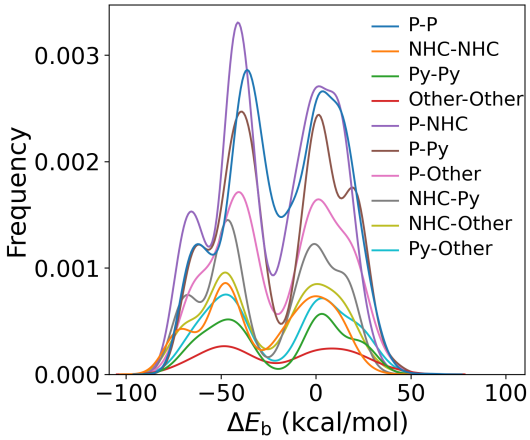


Fig. S2: Kernel density estimation showing distribution of relative binding energies by ligand pair $(L_i - L_j)$. Note that $L_i - L_j$ is considered the same as $L_j - L_i$.

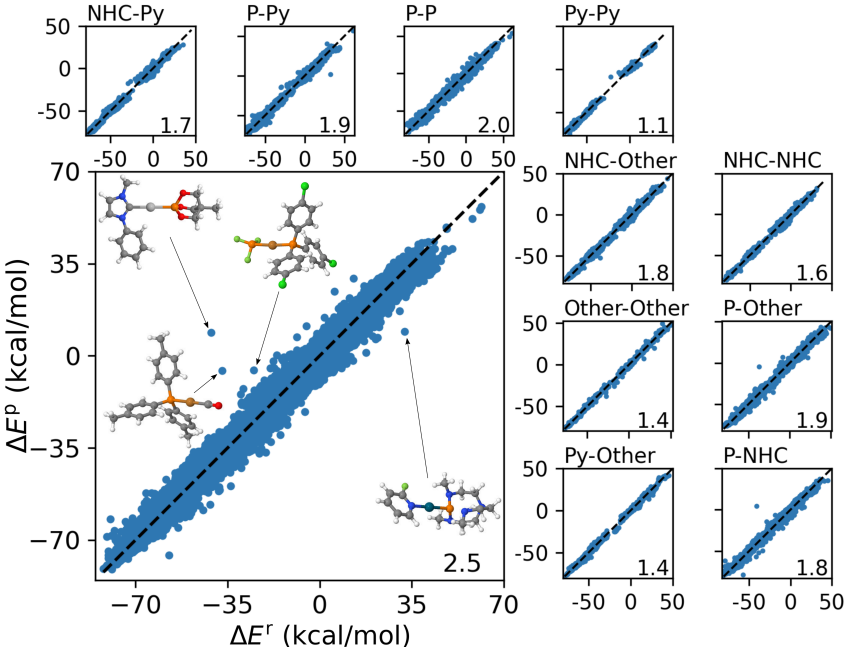


Fig. S3: HIP-predicted relative binding energies against reference energies for full DB1 (bottom left) and subsets based on ligand type combinations (surroundings). Insets display complexes that deviate most (MAE > 18 kcal/mol). Atom colors are gray, blue, white, green, red, orange, silver, dark cyan, and dark orange for C, N, H, Cl, O, P, Ag, Pd, and Cu, respectively.

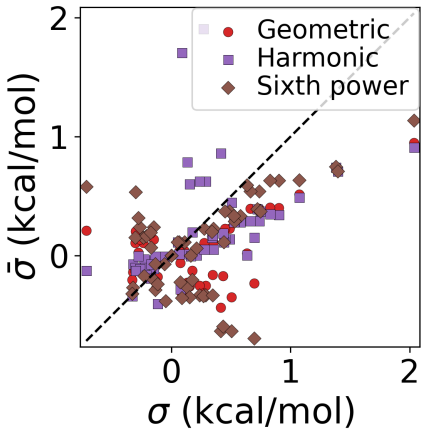


Fig. S4: Test of different combination rules for experimentally obtained substituent parameters describing reactions of various chemistries (see main text). Substituent constants were obtained by combining published Hammett constants according to the rules in section S2.

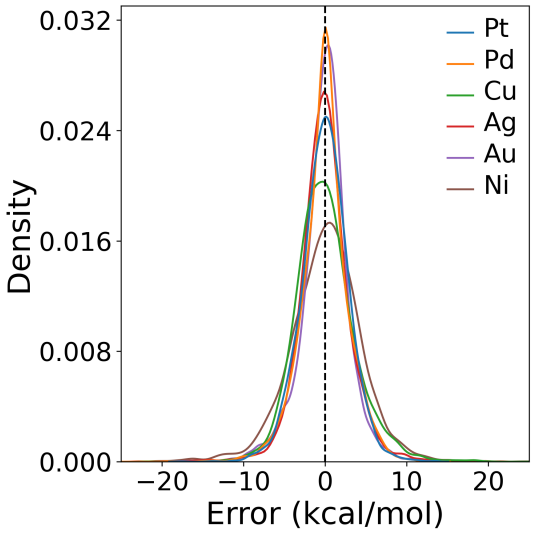


Fig. S5: Error distribution per metal from HIP for DB1.

S4 Supplementary tables

Table S1: Performance of the HIP model fit on DB2.

Reaction step	MAE (kcal/mol)
Oxidation	3.46
Transmetalation	0.88
Reductive elimination	3.31

Table S2: Details of learning curves of KRR direct learning and Δ-ML models using geometry-based representations and performance curve of cHIP baseline model, showing mean absolute error (MAE) ± standard deviation.

Method	**Representation**	**Kernel**	**Training Set Size**	**MAE (kcal/mol)**
Machine learning 7k DFT data	CM	Laplacian	250	8.03 ± 0.23
			500	5.78 ± 0.10
			1000	4.55 ± 0.06
			2000	3.74 ± 0.05
			4000	3.13 ± 0.03
			6000	2.77 ± 0.02
	MBDF	Laplacian	250	5.08 ± 0.17
			500	4.45 ± 0.15
			1000	3.67 ± 0.04
			2000	3.16 ± 0.05
			4000	2.71 ± 0.03
			6000	2.31 ± 0.01
	BoB	Laplacian	250	4.52 ± 0.10
			500	3.93 ± 0.06
			1000	3.53 ± 0.03
			2000	3.11 ± 0.03
			4000	2.68 ± 0.05
			6000	2.41 ± 0.01
	SLATM	Gaussian	250	3.81 ± 0.04
			500	3.95 ± 0.03
			1000	3.23 ± 0.03
			2000	2.99 ± 0.03
			4000	2.81 ± 0.02
			6000	2.69 ± 0.00

Method	Representation	Kernel	Training set size	MAE (kcal/mol)
cHIP on DB1	N/A	N/A	625	67.78 ± 41.89
			1250	4.18 ± 0.27
			2500	4.11 ± 0.31
			5000	4.08 ± 0.31
			10000	4.07 ± 0.31
			20000	4.07 ± 0.31
$\Delta_{\text{Hammett}}^{\text{DFT+ML}}$ on DB1	MBDF	Laplacian	625	9.40 ± 0.22
			1250	2.82 ± 0.04
			2500	2.21 ± 0.02
			5000	1.62 ± 0.02
			10000	1.18 ± 0.01
			20000	0.94 ± 0.00

Table S3: Predicted oxidative addition binding free energies, cHIP parameters, and prices of the 30 cheapest newly discovered catalysts. All energy values are in kcal/mol. Prices are relative to that of the cheapest catalyst as per www.sigmaaldrich.com/CA, retrieved on March 8, 2024.

Metal	**Ligand**$_i$	**Ligand**$_j$	σ_i	σ_j	σ_{ij}	ρ	$\mathbf{\Delta E_{0m}}$	$\mathbf{\Delta E_{oxi}}$	**Price**
Ni	4	71	19.95	14.87	34.82	0.86	-62.16	-32.17	1.00
Pd	95	68	-4.53	0.04	-4.49	0.58	-29.44	-32.05	1.50
Pd	83	68	-0.18	0.04	-0.14	0.58	-29.44	-29.52	1.50
Pd	71	68	14.87	0.04	14.91	0.58	-29.44	-20.81	1.50
Pd	83	95	-0.18	-4.53	-4.71	0.58	-29.44	-32.17	1.50
Pd	71	95	14.87	-4.53	10.33	0.58	-29.44	-23.46	1.50
Pd	83	71	-0.18	14.87	14.69	0.58	-29.44	-20.93	1.51
Pd	71	94	14.87	-8.84	6.03	0.58	-29.44	-25.95	1.51
Pd	71	2	14.87	-9.07	5.79	0.58	-29.44	-26.09	1.59
Pd	82	68	-1.34	0.04	-1.30	0.58	-29.44	-30.20	1.61
Pd	82	95	-1.34	-4.53	-5.88	0.58	-29.44	-32.85	1.61
Pd	82	83	-1.34	-0.18	-1.52	0.58	-29.44	-30.33	1.61
Pd	82	71	-1.34	14.87	13.52	0.58	-29.44	-21.61	1.61
Pd	9	68	-0.11	0.04	-0.07	0.58	-29.44	-29.49	1.78
Pd	9	95	-0.11	-4.53	-4.65	0.58	-29.44	-32.14	1.79
Pd	9	83	-0.11	-0.18	-0.29	0.58	-29.44	-29.61	1.79
Pd	9	71	-0.11	14.87	14.75	0.58	-29.44	-20.90	1.79
Pd	9	82	-0.11	-1.34	-1.46	0.58	-29.44	-30.29	1.89
Pd	8	68	-2.99	0.04	-2.95	0.58	-29.44	-31.15	1.91
Pd	8	95	-2.99	-4.53	-7.52	0.58	-29.44	-33.80	1.91
Pd	8	83	-2.99	-0.18	-3.16	0.58	-29.44	-31.28	1.91
Pd	8	71	-2.99	14.87	11.88	0.58	-29.44	-22.56	1.92
Pd	8	82	-2.99	-1.34	-4.33	0.58	-29.44	-31.96	2.02
Pd	9	8	-0.11	-2.99	-3.10	0.58	-29.44	-31.24	2.20
Pd	91	68	0.31	0.04	0.35	0.58	-29.44	-29.24	2.46
Pd	91	95	0.31	-4.53	-4.22	0.58	-29.44	-31.89	2.46
Pd	83	91	-0.18	0.31	0.14	0.58	-29.44	-29.37	2.46
Pd	91	71	0.31	14.87	15.18	0.58	-29.44	-20.65	2.46
Pd	4	68	19.95	0.04	19.99	0.58	-29.44	-17.86	2.49
Pd	4	95	19.95	-4.53	15.42	0.58	-29.44	-20.51	2.49

Table S4: Constants of individual ligands (kcal/mol) from cHIP on DB1.

Ligand	σ	**Ligand**	σ	**Ligand**	σ
0	21.6	37	-7.59	64	-0.41
10	-2.41	38	-7.27	65	-2.48
11	3.76	39	-13.08	66	7.14
12	2.16	3	1.25	67	22.81
13	-2.06	40	-13.76	68	8.08
14	-0.18	41	-8.37	69	1.92
15	8.69	42	-11.82	6	-0.15
16	28.05	43	-5.83	70	0.5
17	-4.68	44	-4.82	71	29.29
18	22.19	45	-11.02	72	-5.32
19	-7.53	46	-6.48	73	10.48
1	19.43	47	-10.55	74	-0.05
20	5.63	48	-10.93	75	1.31
21	0.24	49	-10.2	76	10.67
22	37.28	4	26.94	77	10.75
23	13.43	50	3.61	78	17.07
24	0.74	51	6.86	79	12.08
25	2.86	52	5.4	7	-3.6
26	2.6	53	3.38	80	18.62
27	-3.4	54	3.31	81	3.26
28	17.26	55	3.43	82	-0.18
29	-3.84	56	1.88	83	0.86
2	-3.27	57	0.34	84	20.22
30	14.9	58	2.8	85	-8.59
31	7.43	59	-0.07	86	-13.94
32	-4.77	5	4.51	87	-11.88
33	-9.95	60	-4.22	88	-4.7
34	-4.66	61	1.78	89	-1.94
35	9.56	62	-0.97	8	40.88
36	-12.98	63	-3.6	90	10.95

Table S5: Constants and offsets (kcal/mol) of metals from cHIP on DB1.

Metal	ρ	$\mathbf{x_0}$
Ag	0.82	18.55
Au	0.75	3.13
Cu	0.92	-6.26
Ni	1.00	-63.82
Pd	1.03	-37.33
Pt	0.85	-43.15

S5 List of ligands

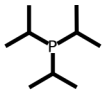
S5.1 Ligands from DB1

The ligands in DB1 were obtained from the publicly-available database published by Meyer *et. al.* [4].

0

1

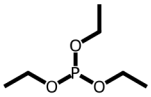
2



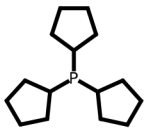
3

4

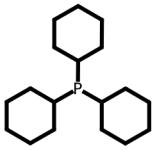
5



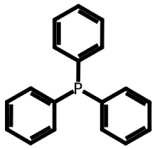
6



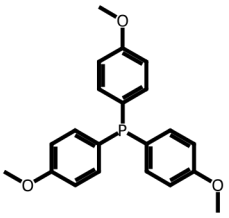
7



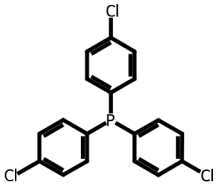
8



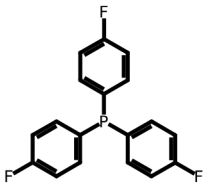
9



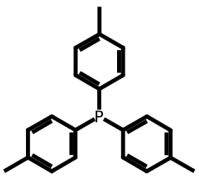
10



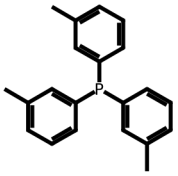
11



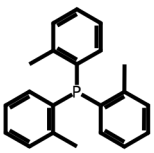
12



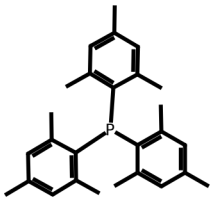
13



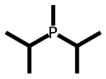
14



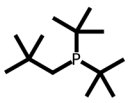
15



16

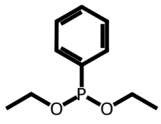


17

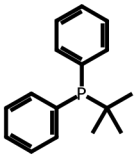


18

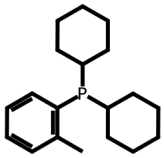
19



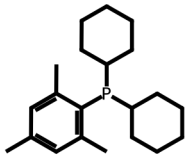
20



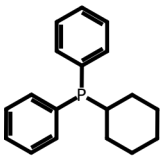
21



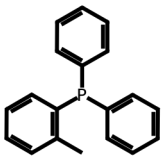
22



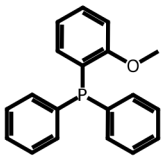
23



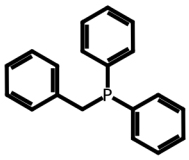
24



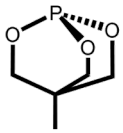
25



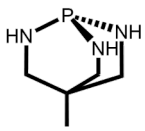
26



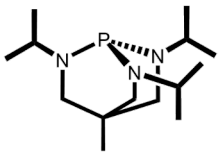
27



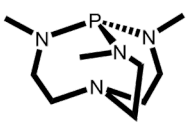
28



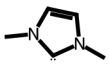
29



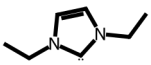
30



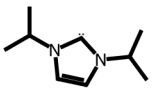
31



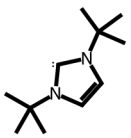
32



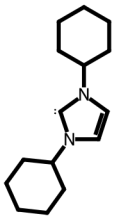
33



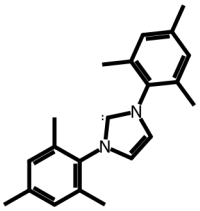
34



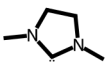
35



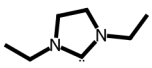
36



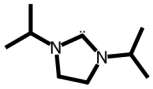
37



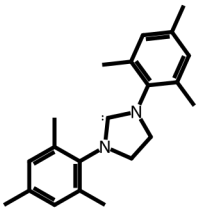
38



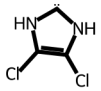
39



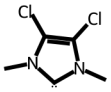
40



41

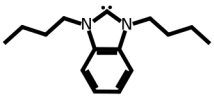


42

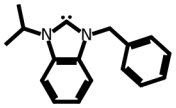


43

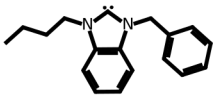
44



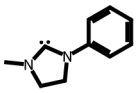
45



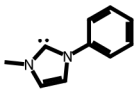
46



47



48



49

50

51

52

53

54

55

56

57

58

59

60

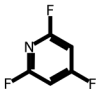
61

62

63

64

65



66

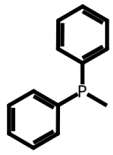
67

68

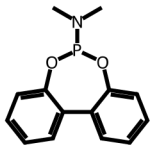
69

70

71

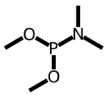


72

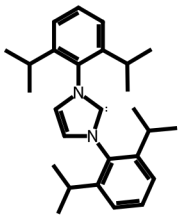


73

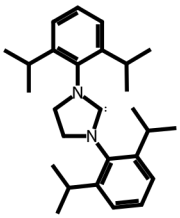
74



75



76

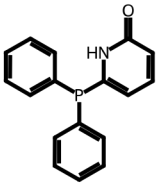


77

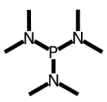
78

79

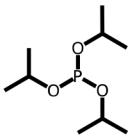
80



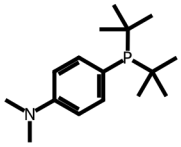
81



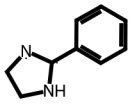
82



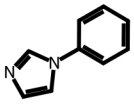
83



84

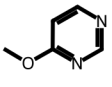


85

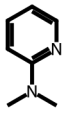


86

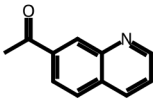
87



88



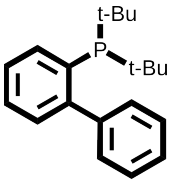
89



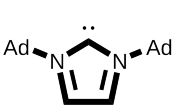
90

S5.2 Ligands from DB2

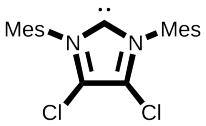
DB2 [5] also contained Ligands no. 2, 4, 8, 9, 41, 68, 71, 81, 82, 83, and 84 from DB1. The remaining ligands are drawn below.



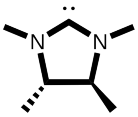
91



92



93



94

95

S6 References

[1] D. Berthelot, Sur le mélange des gaz, Compt. Rendus **126**, 15 (1898).
[2] M. Waldman and A. T. Hagler, New combining rules for rare gas van der waals parameters, Journal of computational chemistry **14**, 1077 (1993).
[3] B. Fender and G. Halsey Jr, Second virial coefficients of argon, krypton, and argon-krypton mixtures at low temperatures, The Journal of Chemical Physics **36**, 1881 (1962).
[4] B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld, and C. Corminboeuf, Machine learning meets volcano plots: computational discovery of cross-coupling catalysts, Chemical science **9**, 7069 (2018).
[5] M. Busch, M. D. Wodrich, and C. Corminboeuf, A generalized picture of C-C cross-coupling, ACS Catalysis **7**, 5643 (2017).