Electronic Supplementary Material (ESI) for Digital Discovery.

Tailoring Phosphine Ligands for Improved C-H Activation: Insights from Δ-Machine Learning

Tianbai Huang,[a] Robert Geitner,[b,*] Alexander Croy,[a,*] Stefanie Gräfe,[a,*]

[a] Institute for Physical Chemistry (IPC) and Abbe Center of Photonics, Friedrich Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany

[b] Institute of Chemistry and Bioengineering, Technical University Ilmenau, Weimarer Str. 32,98693 Ilmenau, Germany

*e-mail: robert.geitner@tu-ilmenau.de, alexander.croy@uni-jena.de, s.graefe@uni-jena.de

Details of atomic properties used in ACs and AC-AIMs

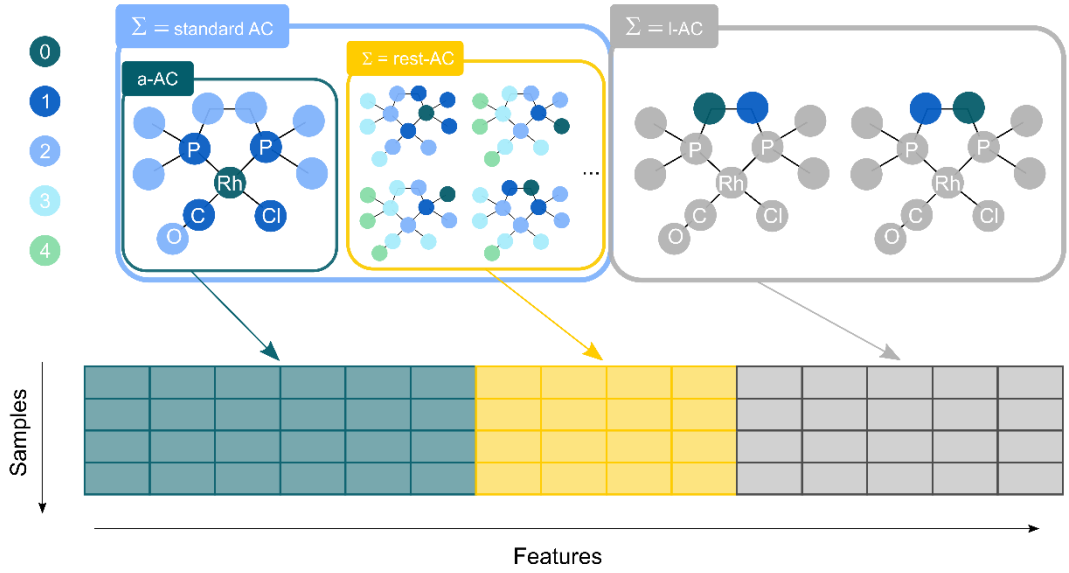

Fig. S1 Schematic of a-ACs, rest-ACs, standard ACs and l-ACs using as a representative model (Hydrogens are omitted for clarity). In the illustration of a-AC, atoms in the same topological distance with respect to the selected atom were represented in the same color. The standard ACs are the sum of a-ACs evaluated for all atoms in the molecule, while rest-ACs are the sum of a-ACs excluded from the selected set of atoms. In order to obtain the detailed description for both local and global properties of the Rh complexes, the a-ACs, rest-ACs and l-ACs are combined to form the feature vector.

Table S1 List of atomic properties A^{P} for molecular Autocorrelation functions (ACs) M^{P}.

DFT-based single-atom properties	Element-specific parameters in GFN2-xTB[1]	AIM-based atomic properties
identity (id) binary ring index ($Ring$) nuclear charge (Z) electron affinity (EA) ionization potential (IP) electronic spatial extent in the atom ($\langle r^2 \rangle$) electronic spatial extent in the cation ($\langle r_+^2 \rangle$) electronic spatial extent in the anion ($\langle r_-^2 \rangle$) vdw-volume of atom (V) vdw-volume of cation (V_+) vdw-volume of anion (V_-) maximum LOL value of the atom ($\gamma_{\max}$) maximum LOL value of the cation ($\gamma_{\max+}$) maximum LOL value of the anion ($\gamma_{\max-}$) distance of maximum LOL value in the atom ($d_{\gamma_{\max}}$) distance of maximum LOL value in the cation ($d_{\gamma_{\max+}}$) distance of maximum LOL value in anion ($d_{\gamma_{\max-}}$)	atomic Hubbard parameter (η_A) charge derivative of Hubbard parameter (Γ_A) exponential scaling parameters ($\alpha_A, Y_A^{\mathrm{eff}}$) anisotropic XC scaling parameters ($f_{\mathrm{XC}}^{\mu_A}, f_{\mathrm{XC}}^{\Theta_A}$) offset radius ($R_0^A$)	atomic charge (chg) integration of the norm of electron density gradient ($\int \lvert\nabla\rho\rvert$) atomic volume ($\mathrm{Vol}$) Lagrangian kinetic energy (Kin) localization index (LI) delocalization index (DI)
	Element-specific shell parameters of the most outer shell in GFN2-xTB	
	polynomial scaling parameters ($k_{A,l}^{\mathrm{poly}}$) shell-specific scaling parameters of the Hubbard parameter (κ_A^l) the constant part of the energy levels (H_A^l) coordination-dependent enhancement factors for energy levels ($H_{CN'_A}^l$) Slater exponents (ζ_l)	

Except for the feature identity (id), binary ring index ($Ring$) and the nuclear charge (Z), all the features in the first column in

Table **S1** were derived from single-point calculations on a single neutral atom, as well as its cationic (+1) and anionic (-1) states, using the DFT approach. All calculations were carried out using the ωB97XD functional[2] and def2SVP basis set.[3] Further analysis of the electron densities was conducted, utilizing the Multiwfn[4] software to generate the atomic properties. The isolated atomic properties used to construct the ACs in our study include the following:

1. **Electron affinity** (EA): This is defined as the energy difference between the neutral atom and its cationic state.
2. **Ionization potential** (IP): This is defined as the energy difference between the neutral atom and its anionic state.
3. **Electronic spatial extent** $\langle r^2 \rangle$ of an atom is defined as follows:

$$\langle r^2 \rangle = \int r^2 \rho(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}\,. \qquad (1.)$$

4. **Vdw-volume** of an atom V is defined as the space bounded iso-surface of the electron density, where $\rho(\boldsymbol{r})$ is 0.001 a.u.:

$$V = \int_{\rho(\boldsymbol{r}) > 0.001\ \text{a.u.}} 1\ \mathrm{d}\boldsymbol{r}\,. \qquad (2.)$$

The above two atomic properties $\langle r^2 \rangle$ and V are two distinct ways to a measure of the spatial distribution of electrons. Furthermore, information of the localized orbital locator (LOL)[5] is also incorporated, which is defined as:

$$\gamma(\boldsymbol{r}) = \frac{t(\boldsymbol{r})}{1 + t(\boldsymbol{r})}, \qquad t(\boldsymbol{r}) = D_0 \frac{\rho_\sigma^{5/3}(\boldsymbol{r})}{\sum_i |\nabla \psi_{i\sigma}(\boldsymbol{r})|^2}\,, \qquad (3.)$$

where the numerator in $t(\boldsymbol{r})$ is proportional to the kinetic energy densities under the local density approximation, whereas the denominator is the non-interacting Kohn-Sham kinetic energy density. The vicinity of the maxima point of LOL value in the atomic system is associated to the atomic shell that accumulate electronic charge, which might play an important role in the bond formation. Therefore, we selected two characteristics from this function, the **maximum of LOL value** $\gamma_{\max}$ and **the distance of the maximum LOL value from the nuclei** $d_{\gamma_{\max}}$ (see Fig. **S2**), as the atomic properties to construct the ACs. Furthermore, the $\langle r^2 \rangle$, V, $\gamma_{\max}$ and $d_{\gamma_{\max}}$ of the cation and anion states of the corresponding elements were also used to construct the ACs.

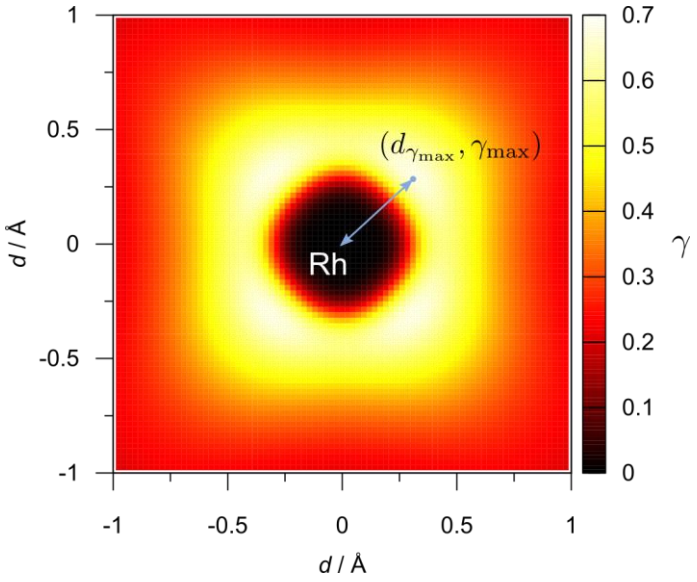

Fig. S2 The illustration of the localized orbital locator (LOL) function γ of Rh atom. The distance between the maximum point of γ and the nucleus is defined as $d_{\gamma_{\max}}$.

In addition to the DFT-based isolated atomic properties, the element-specific xTB parameters were also used as the atomic properties (second column in Table **S1**) in our study to construct the ACs. All the element-specific parameters are well-detailed in the original GFN2-xTB paper[1] as well as the corresponding supporting information.

The atomic properties calculated using AIM theory[6, 7] were also employed to construct the feature sets AIM-ACs in our study. Properties include the atomic charge $\mathrm{chg}(\Omega)$, atomic volume $\mathrm{Vol}(\Omega)$, the integral of gradient norm of the electron density within the atomic basin $\int|\nabla\rho|(\Omega)$, the Lagrangian kinetic energy $\mathrm{Kin}(\Omega)$, the localization index $\mathrm{LI}(\Omega)$ and the delocalization index $\mathrm{DI}(\Omega)$. For a comprehensive understanding of these AIM-based atomic properties, readers are referred to [7] and the user manual of Multiwfn[4].

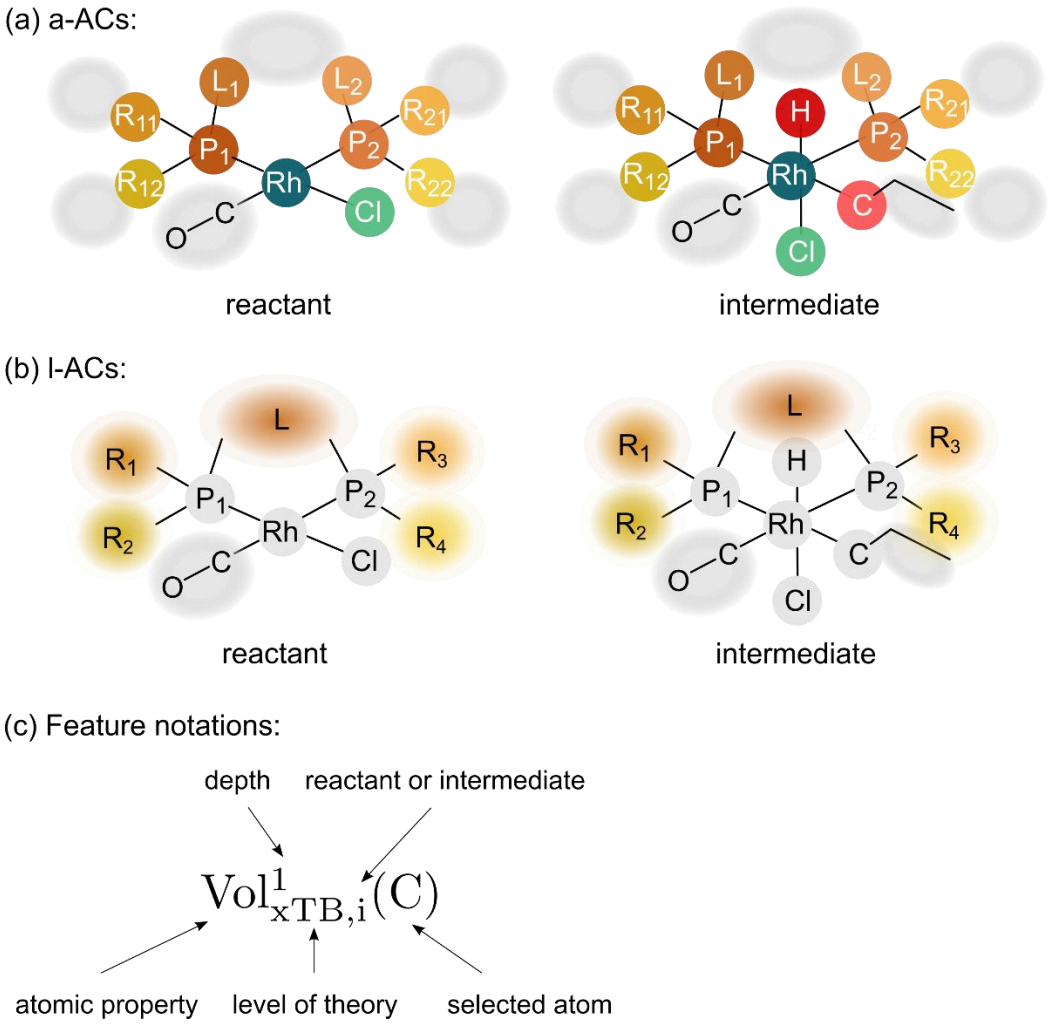

Fig. S3 (a) Selected atoms in reactant and intermediate for calculating the a-ACs. (b) Building blocks for the calculation of l-ACs. (c) Notation of the AC and AIM-AC features.

Table S2 Dimension and the feature dependence of AC-, AIM-AC-, SOAP-based features before and after dimensionality reduction, as well as root-mean-squared error (RMSE) evaluated on the test set (validation set in parentheses), R^2 evaluated on the test set and hyperparameters of optimal $\Delta^{D'}_{x'}$-models and Δ^{D}_{x}-models. Both types of models were trained in the framework of artificial neural network (ANN) using the features after the dimensionality reduction.

$\Delta E'_{\mathrm{xTB}} \rightarrow \Delta E'_{\mathrm{DFT}}$		number of features before / after	RMSE / kJ mol^{-1}	R^2	array of neurons / embedding dropout rate / linear dropout rate / number of epochs	$\Delta E_{\mathrm{xTB}} \rightarrow \Delta E_{\mathrm{DFT}}$		number of features before / after	RMSE / kJ mol^{-1}	R^2	array of neurons / embedding dropout rate / linear dropout rate / number of epochs
ACs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{r}}, \boldsymbol{R}_{\mathrm{i}})$	1	2121 / 526	16.0 (15.8)	0.803	(1014, 1223) / 0.115 / 0.236 / 29	ACs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{r}}, \boldsymbol{R}_{\mathrm{i}})$	1	2121 / 534	12.7 (13.2)	0.723	(1497, 701) / 0.489 / 0.125 / 24
	3	4241 / 476	14.8 (14.8)	0.832	(1122, 391, 312, 327) / 0.249 / 0.057 / 24		3	4241 / 585	13.3 (12.5)	0.699	(450, 234, 193, 206) / 0.332 / 0.364 / 28
	5	6361 / 330	14.6 (14.7)	0.837	(486, 354, 430, 411) / 0.424 / 0.063 / 20		5	6361 / 445	13.1 (12.9)	0.705	(552, 494, 407, 348) / 0.729/ 0.204 / 24
	7	8481 / 328	14.2 (14.7)	0.845	(1301, 221, 202) / 0.489 / 0.044 / 30		7	8481 / 377	13.2 (12.8)	0.698	(928, 150) / 0.061 / 0.317 / 26
	9	10601 / 336	14.6 (15.1)	0.834	(947, 144, 85, 95, 104) / 0.731 / 0.008 / 27		9	10601 / 379	12.8 (12.6)	0.706	(368, 278, 284) / 0.483 / 0.068 / 28
AIM-ACs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	1	637 / 440	11.1 (10.4)	0.905	(929, 322) / 0.087 / 0.127 / 26	AIM-ACs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	1	637 / 577	10.6 (9.6)	0.809	(610, 234, 194) / 0.049 / 0.182 / 22
	3	1273 / 364	10.6 (10.6)	0.913	(1042, 367, 150, 121) / 0.651 / 0.002 / 24		3	1273 / 668	10.6 (9.7)	0.809	(964, 455, 337) / 0.692 / 0.260 / 25
	5	1909 / 296	10.7 (10.6)	0.911	(1327, 796) / 0.306 / 0.000 / 29		5	1909 / 590	10.7 (9.8)	0.804	(1181, 1721, 695, 306) / 0.527 / 0.272 / 24
	7	2545 / 257	10.9 (11.0)	0.909	(1422, 767) / 0.778 / 0.114 / 30		7	2545 / 522	10.6 (10.0)	0.808	(943, 508) / 0.657 / 0.080 / 23
	9	3181 / 249	11.1 (10.5)	0.905	(1230, 476, 281, 131, 119) / 0.492 / 0.112 / 27		9	3181 / 507	10.4 (10.3)	0.816	(721, 734, 787, 843, 692) / 0.421 / 0.052 / 24
SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$		6073 / 463	13.0 (11.4)	0.871	(1280, 1095, 442, 389, 480) / 0.585 / 0.309 / 30	SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$		9109 / 672	10.3 (8.7)	0.819	(566, 302) / 0.418 / 0.247 / 25
-						xTB-SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$		6073 / 774	11.1 (9.6)	0.789	(423, 209, 131) / 0.221 / 0.188 / 27

Table S3 Comparison of the performances of artificial neural network (ANN), XGBoost, CatBoost models trained on ACs ($d_{\max} = 7$), AIM-ACs ($d_{\max} = 3$) and SOAPs for $\Delta E'_{\mathrm{DFT}}$, as well as the models trained on ACs ($d_{\max} = 1$), AIM-ACs ($d_{\max} = 9$), SOAPs and xTB-SOAPs for ΔE_{DFT}.

$\Delta E'_{\mathrm{xTB}} \rightarrow \Delta E'_{\mathrm{DFT}}$	ANN		XGBoost		CatBoost		$\Delta E_{\mathrm{xTB}} \rightarrow \Delta E_{\mathrm{DFT}}$	ANN		XGBoost		CatBoost	
	RMSE / kJ mol^{-1}	R^2	RMSE / kJ mol^{-1}	R^2	RMSE / kJ mol^{-1}	R^2		RMSE / kJ mol^{-1}	R^2	RMSE / kJ mol^{-1}	R^2	RMSE / kJ mol^{-1}	R^2
ACs ($d_{\max} = 7$) $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{r}}, \boldsymbol{R}_{\mathrm{i}})$	14.2 (14.7)	0.845	14.5 (15.7)	0.839	13.6 (15.5)	0.857	ACs ($d_{\max} = 1$) $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{r}}, \boldsymbol{R}_{\mathrm{i}})$	12.7 (13.2)	0.723	13.3 (13.2)	0.698	13.3 (13.4)	0.697
AIM-ACs ($d_{\max} = 3$) $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	10.6 (10.6)	0.913	12.4 (12.8)	0.882	12.5 (12.8)	0.880	AIM-ACs ($d_{\max} = 9$) $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	10.4 (10.3)	0.816	11.0 (11.6)	0.793	11.0 (11.5)	0.791
SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	13.0 (11.4)	0.871	13.0 (13.2)	0.869	13.2 (13.1)	0.866	SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{DFT,r}}, \boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	10.3 (8.7)	0.819	11.5 (11.4)	0.774	11.5 (11.4)	0.773
-							xTB-SOAPs $\boldsymbol{\sigma}(\boldsymbol{R}_{\mathrm{xTB,r}}, \boldsymbol{R}_{\mathrm{xTB,i}})$	11.1 (9.6)	0.789	11.7 (11.7)	0.766	11.9 (11.6)	0.756

Details of Smooth overlap atomic positions (SOAPs) for describing the reactant-intermediate pairs

SOAPs[8] are a type of position-based features that describe the atomic environment of the selected atoms. The SOAP feature describing a molecule could be generated by combining atomic SOAPs of a selected set of atoms, as illustrated in Fig. **S4**. In our study, the SOAP features of a single atom α is represented by a vector, whose entries are defined as:

$$p_{nn'l}^{Z_1Z_2}(\alpha) = \pi \sqrt{\frac{8}{2l+1}} \sum_m \left(c_{nlm}^{Z_1}\right)^* \left(c_{n'lm}^{Z_2}\right), \tag{4.}$$

where Z_1 and Z_2 denote the atomic species, n and n' are indices for different radial basis functions and l is the angular degree of the spherical harmonics. The coefficient c_{nlm}^{Z} is the measure of the atomic density of kind Z evaluated on the spherical harmonic $Y_{lm}(\theta, \varphi)$ and radial basis $g_n(r)$ functions,

$$c_{nlm}^{Z}(\alpha) = \iiint_{\mathbb{R}^3} \mathrm{d}V \; g_n(r) Y_{lm}(\theta, \varphi) \, \rho^Z(\boldsymbol{r}), \tag{5.}$$

where $\rho^Z(\boldsymbol{r})$ is the gaussian smoothed atomic density of kind Z around the specified atom α, defined as

$$\rho^Z(\boldsymbol{r}) = \sum_j^{n_Z} \mathrm{e}^{-\frac{|\boldsymbol{r}-\boldsymbol{R}_j|^2}{2\sigma^2}}, \tag{6.}$$

where $\boldsymbol{r}$ is the relative position with respect to atom α and $\boldsymbol{R}_j$ is the relative position of environmental atom j with respect to atom α. n_Z is the total number of atoms of kind Z in the molecule.

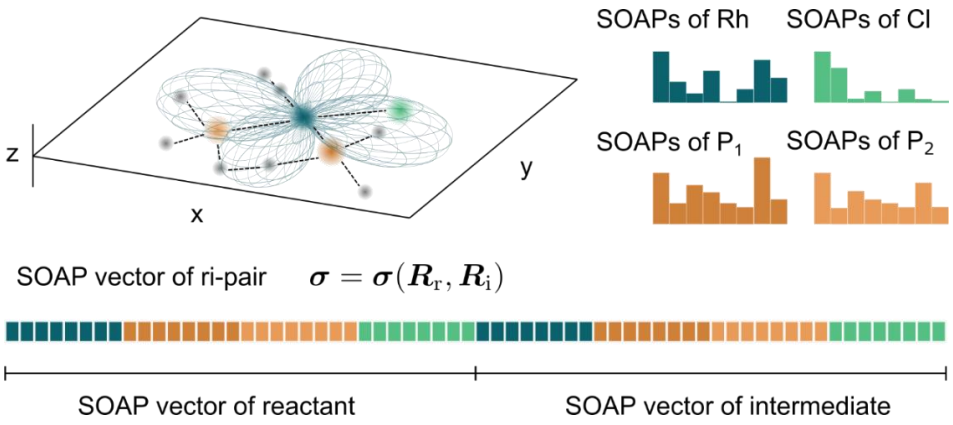

Fig. S4 Graphical illustration of the SOAP vector of a reactant-intermediate pair. Hereby the feature vector $\boldsymbol{\sigma}$ depending on $\boldsymbol{R}_\mathrm{r}$ and $\boldsymbol{R}_\mathrm{i}$ is used as an illustrative example.

The Building Blocks of Ligands

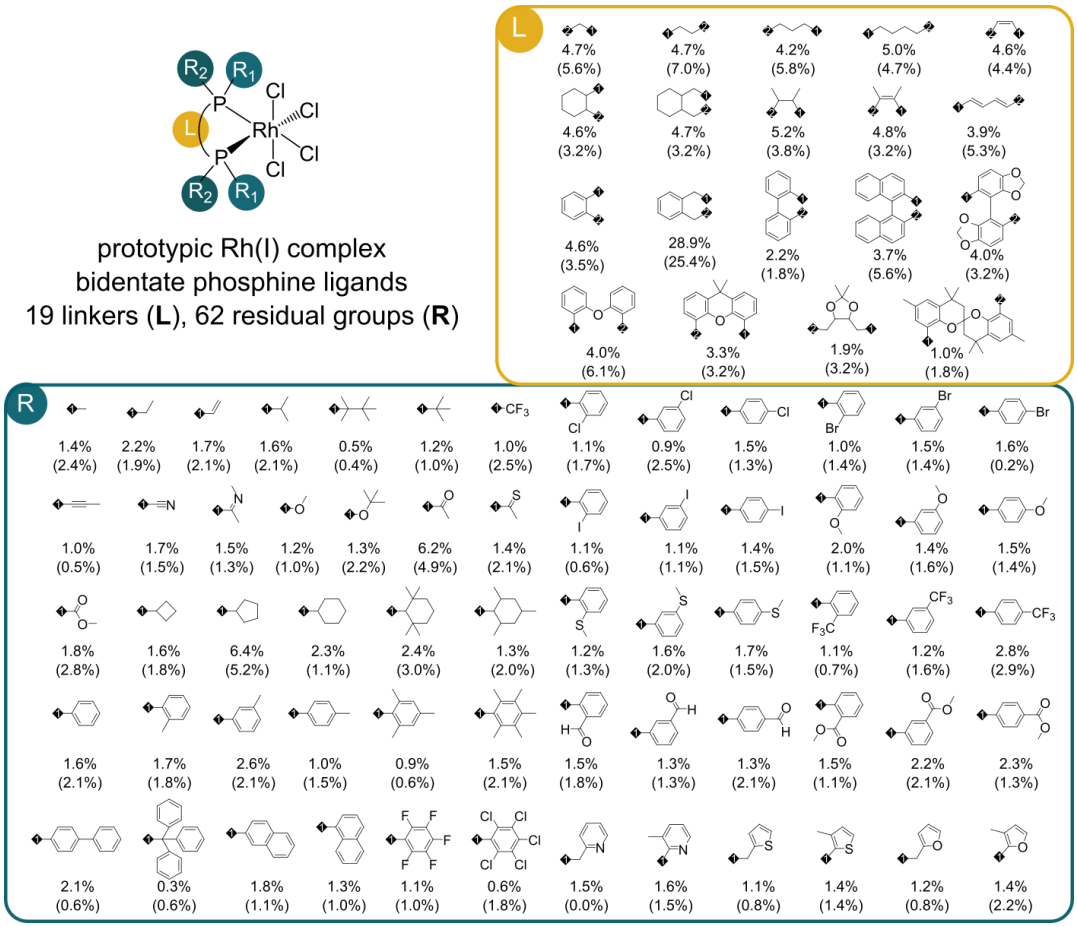

Fig. S5 Selected residual groups (**R**) and linkers (**L**) that define the region of chemical space associated with bidentate phosphine ligands. The chemical space of the ligands is constrained so that at most two different residual groups R_1 and R_2 are allowed in one ligand with each phosphorous atom possessing both R_1 and R_2. The distribution of building blocks in training and test set (in parenthesis) are also provided. The high fraction of the *o*-xylene unit in the linker is due to the inclusion of the related conformers. **R** and **L** were selected in such a way to ensure a realistic synthetic pathway to the optimized bisphosphine ligands.

Results of $\Delta_{\mathrm{x}}^{\mathrm{D}}$-model trained on ACs with $d_{\max} = 1$

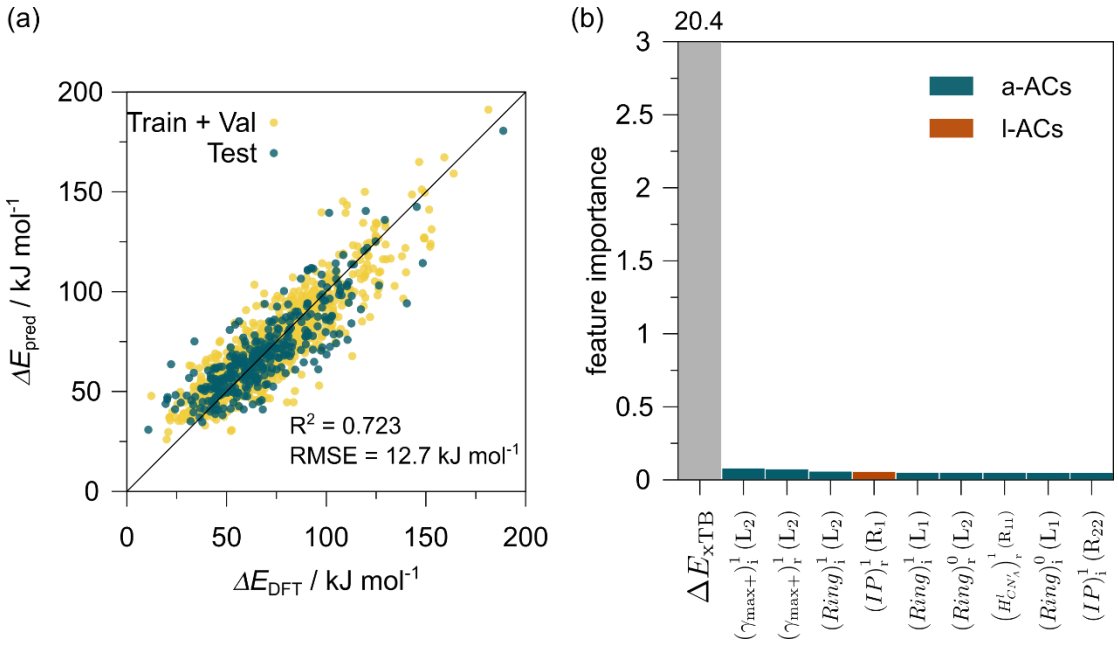

Fig. S6 (a) Parity plots between the prediction values and the true (DFT) value of the best Δ_x^{D}-models trained on ACs ($d_{\max} = 1$). (b) The corresponding feature importance analysis.

Detailed description of the two-step high-throughput screening using $\Delta_{\mathrm{x}}^{\mathrm{D}}$-model trained on xTB-SOAPs

The overall process is illustrated in

Fig. **S7** (a), encompassing a preliminary screening without conformer search and a subsequent refined screening with conformer search. In the preliminary screening, 27,832 distinctive asymmetric prototypic complexes Rh(PLP)$(Cl)_4$ were generated by combinatorically selecting different **L** linkers and **R** residues as the ligand building blocks. The corresponding 4-coordinated reactants and 6-coordinated intermediates were then derived through the substitution of the Cl ions. Features, including SOAPs and the baseline value ΔE_{xTB} were evaluated using the xTB-equilibrium structures of reactants and intermediates. Subsequently, the ΔE_{DFT} of these 27,832 reactant-intermediate pairs were estimated by ΔE_{ML}, derived from the trained $\Delta_{\mathrm{x}}^{\mathrm{D}}$-model using the SOAPs and ΔE_{xTB} as inputs. It is worthy noting that the reactant and intermediate geometries obtained in the preliminary screening might be energetically far away from the global minimum in their conformer space, the population of which is insignificant in the equilibrated reaction system. Therefore, a subsequent screening procedure with a

conformer search was conducted on a selected set of 60 reactant-intermediate pairs with the lowest predicted reaction energies (as highlighted in red in
Fig. **S7** (b)). In the refined screening, the corresponding prototypes of those 60 reactant-intermediate pairs underwent a conformer search using CREST. A set of the conformers of reactant and intermediate was generated by substituting Cl ions of the prototypic conformers of the same complex. The features, including SOAPs and ΔE_{xTB} of the corresponding reactant-intermediate pair, were obtained from the intermediate and reactant with the lowest energy at their xTB-equilibrium geometries in the conformer set. The ΔE_{DFT} of the reactant-intermediate pairs at their optimal conformational structure were estimated once again by the ΔE_{ML}, as highlighted in red in
Fig. **S7** (c).

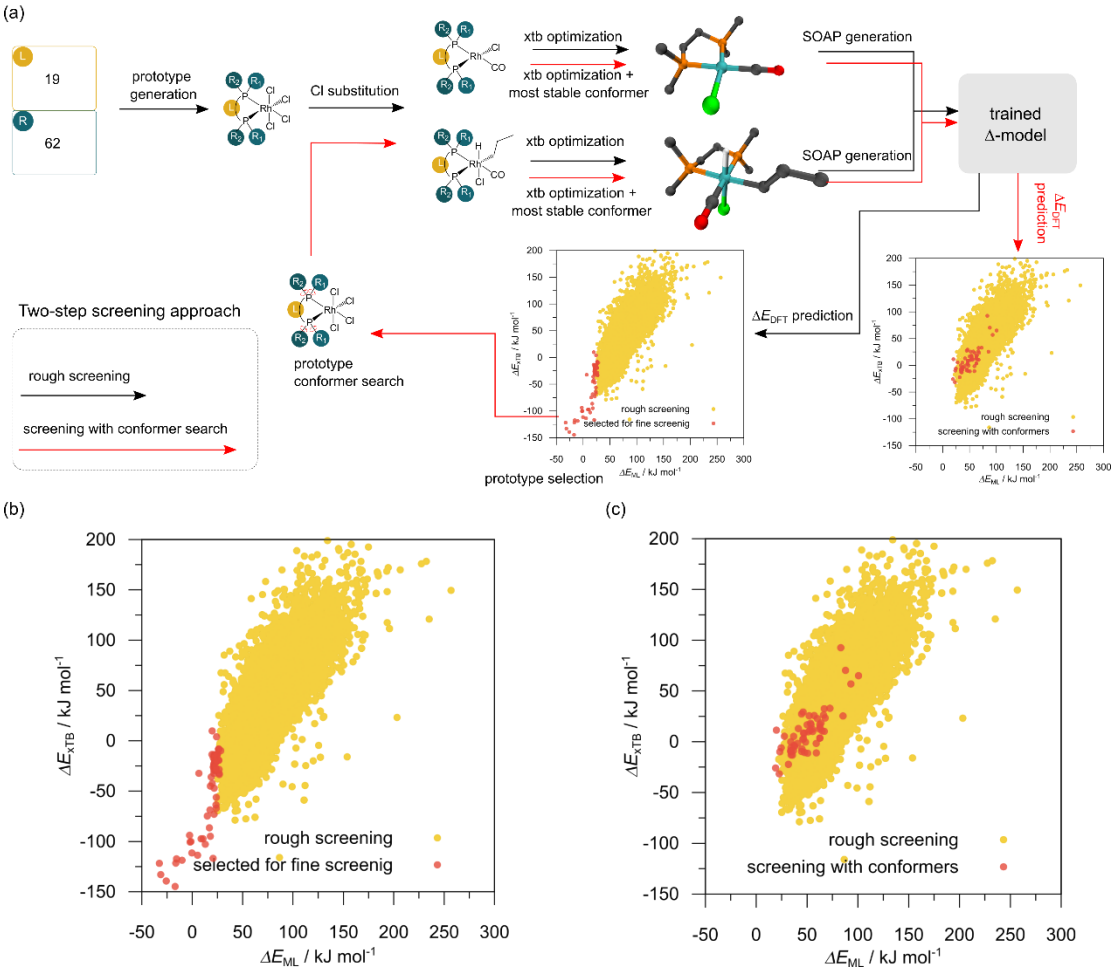

Fig. S7. (a) The illustration of two-step screening procedure. In the rough screening step, the corresponding reactant-intermediate pairs were obtained by substituting the Cl ions from the prototypes featuring different ligands. SOAPs and ΔE_{xTB} of the reactant-intermediate pairs were obtained from the xTB-equilibrium structures of the reactants and intermediates. Utilizing these features as input of the trained $\Delta_{\mathrm{x}}^{\mathrm{D}}$-model, the ΔE_{ML} were obtained as an estimation of the ΔE_{DFT} of the reactant-intermediate pairs with varying bidentate phosphine ligands. The fine screening procedure was employed on the selected set of reactant-intermediate pairs with low ΔE_{ML}. Except for an additional conformer search on the corresponding prototypic molecule Rh(PLP)(Cl)$_4$ and the selection of optimal structures for reactants and intermediates, other procedures remained identical as in the rough screening procedure. (b) Scatter plot of the ΔE_{ML} versus ΔE_{xTB} evaluated on 27,832 data points obtained from the rough screening process. Data points highlighted in red were selected for the fine screening. (c) Scatter plot (highlighted in red) of the ΔE_{ML} versus ΔE_{xTB} evaluated on 60 data points obtained from the fine screening process.

Statistical Analysis of Geometric and Electronic Factors for DFT-optimized Equilibrium Structures

The analysis was done using R (4.3.1).[9] First, the DFT-calculated equilibrium structures for the complexes [Rh(PLP)(CO)(Cl)] were read in including the xyz coordinates, the Mullikan charges q, the vibrational frequencies, their IR intensity and Raman activity, the dipole moment μ as well as the HOMO and LUMO energies E_{HOMO} and E_{LUMO}. From the xyz coordinates the two Rh-P distances $d_{\mathrm{Rh,P1}}$ and $d_{\mathrm{Rh,P2}}$ were calculated, while the vibrational frequency $\nu(\mathrm{CO})$ as well as the IR intensity $I_{\mathrm{IR}}(\mathrm{CO})$ and Raman activity $\alpha_{\mathrm{Raman}}(\mathrm{CO})$ for the CO vibration were extracted from the vibrational data. The HOMO-LUMO energy difference $\Delta E_{\mathrm{HOMO-LUMO}}$ was derived from the two frontier molecular orbital energies. Using the R packages dplyr (1.1.2),[10] stringr (1.5.1),[11] htmlTable (2.4.2),[12] gtools (3.9.5)[13] and reticulate (1.34.0)[14] as well as the python package morfeus (0.7.2),[15] the cone angle Θ, the bite angle β, the (buried) Sterimol parameters (buried) $B1$, $B5$ and L, the solvent accessible surface area and volume $\mathrm{SAS}A$ and $\mathrm{SAS}V$ for the entire complex and the volume for the Rh atom $\mathrm{SAS}V_{\mathrm{Rh}}$, the solid angle cone angle Ω, the %buried volume $\%V\mathrm{Bur}$, the total buried volume $V\mathrm{Bur}$, the distal volume $\mathrm{Dis}V$, the eight octant volumes $\mathrm{O}V_{1-8}$, the two hemisphere volumes $\mathrm{H}V_{\uparrow}$ and $\mathrm{H}V_{\downarrow}$ (all centered on the Rh atom, with the z-axis defined as the direction from the Rh atom to the mean position of the two P atoms and using a sphere radius of 3.5 Å), the pyramidalization P and pyramidalization angle α for the two P atoms were calculated. Furthermore, the xyz coordinates were used as input for xTB calculations to access the Fukui parameters f^{+}, f^{-} and f^{0} for Rh and the two P atoms as well as the ionization potential IP, electron affinity EA, electrophilicity ω, nucleophilicity N, electrofugality ν_{EleFuc}, and nucleofugality ν_{NucFuc} of the entire complex. In total 59 structural factors were derived.

To judge if a significant difference between the original dataset with 1743 data points and the ten newly proposed structures was present a two-sided t-test was performed for each factor. A significant difference was assumed when the p-value of the t-test was below 0.05, which is equivalent to a certainty of 95 %. This led to the identification of 20 significantly different geometric and electronic factors between the two datasets (see Table **S4**).

Table S4. Geometric and electronic factors from the original and the newly proposed bisphosphine set derived for the corresponding complexes [Rh(PLP)(CO)(Cl)]. In addition to the p-value of the t-test the mean (μ) and the standard deviation (σ) for both datasets are also given. The significance is judged on a threshold of 0.05.

Factor	p-value	μ_{ori}	μ_{new}	σ_{ori}	σ_{new}	Significant?
Buried $B5$ / Å	10^{-13}	7.33	7.79	0.63	0.08	TRUE
q_{Rh} / au	10^{-4}	-0.35	-0.32	0.05	0.02	TRUE
q_{P1} / au	10^{-4}	-0.41	-0.45	0.07	0.03	TRUE
f_{P2}^{-}	10^{-3}	-0.04	-0.03	0.02	0.01	TRUE
μ / D	10^{-3}	8.33	6.76	1.72	1.02	TRUE
$I_{IR}(CO)$ / km mol^{-1}	10^{-3}	863	809	75	37	TRUE
q_{P2} / au	10^{-3}	0.20	0.22	0.03	0.01	TRUE
α_{P2} / °	10^{-3}	6.83	7.19	0.47	0.25	TRUE
P_{P2}	10^{-3}	0.88	0.91	0.04	0.02	TRUE
P_{P1}	10^{-3}	0.89	0.92	0.04	0.02	TRUE
$B5$ / Å	10^{-3}	8.37	9.62	1.43	0.94	TRUE
α_{P1} / °	10^{-3}	7.01	7.31	0.42	0.25	TRUE
$\alpha_{Raman}(CO)$ / Å^4 amu^{-1}	0.01	60	52	69	6	TRUE
$HV_{\downarrow}$ / Å^3	0.01	79	75	5	3	TRUE
$d_{Rh,P2}$ / Å	0.01	1.87	1.73	0.24	0.15	TRUE
$\%V\mathrm{Bur}$	0.02	0.52	0.49	0.05	0.02	TRUE
$V\mathrm{Bur}$ / Å^3	0.02	93	89	10	4	TRUE
OV_4 / Å^3	0.04	3.45	2.33	2.39	1.46	TRUE
$\nu(CO)$ / cm^{-1}	0.04	2161	2171	19	13	TRUE
OV_5 / Å^3	0.04	19.7	18.0	2.3	2.3	TRUE
q_{Cl} / au	0.05	0.38	0.33	0.12	0.07	FALSE
f_{P2}^{0}	0.05	-0.06	-0.05	0.03	0.01	FALSE
ν_{EleFuc} / eV	0.05	0.14	0.08	0.16	0.08	FALSE
Ω / °	0.06	186.3	181.4	14.9	7.3	FALSE
Ω / G	0.07	0.53	0.51	0.06	0.03	FALSE
β / °	0.07	90.27	92.97	8.03	4.12	FALSE
$SASA$ / Å^2	0.07	774	834	145	91	FALSE
q_C / au	0.08	0.33	0.29	0.12	0.06	FALSE
f_{Rh}^{0}	0.09	-0.05	-0.05	0.02	0.01	FALSE
OV_8/ Å^3	0.11	19.5	17.9	2.4	2.8	FALSE
Buried $B1$ / Å	0.12	4.90	4.60	0.70	0.56	FALSE
E_{HOMO} / kJ mol^{-1}	0.17	-718	-734	38	33	FALSE
OV_2 / Å^3	0.20	3.52	4.57	2.42	2.39	FALSE

f_{P2}^{+}	0.20	-0.07	-0.06	0.04	0.03	FALSE
$\Delta E_{HOMO-LUMO}$ / kJ mol^{-1}	0.22	734	750	42	37	FALSE
f_{P1}^{0}	0.22	-0.08	-0.07	0.03	0.02	FALSE
$SASV_{Rh}$ / Å^3	0.22	3.73	4.86	3.37	2.71	FALSE
f_{Rh}^{-}	0.22	-0.09	-0.09	0.02	0.02	FALSE
f_{P1}^{+}	0.22	-0.08	-0.07	0.04	0.03	FALSE
$SASV$ / Å^3	0.25	1453	1531	311	195	FALSE
OV_1 / Å^3	0.28	3.47	2.92	2.36	1.48	FALSE
f_{Rh}^{+}	0.29	-0.02	-0.01	0.03	0.01	FALSE
EA / eV	0.31	1.91	2.07	0.64	0.45	FALSE
ν_{NucFuc} / eV	0.38	9.74	9.92	0.81	0.60	FALSE
ω	0.44	2.05	2.15	0.51	0.38	FALSE
f_{P1}^{-}	0.46	-0.07	-0.07	0.02	0.01	FALSE
$DisV$ / Å^3	0.48	482	504	153	92	FALSE
IP / eV	0.49	7.69	7.77	0.44	0.33	FALSE
N / eV	0.49	-7.69	-7.77	0.44	0.33	FALSE
$d_{Rh,P1}$ / Å	0.54	1.82	1.84	0.23	0.10	FALSE
$B1$ / Å	0.55	5.19	5.02	0.92	0.88	FALSE
Θ / °	0.62	224	227	15	16	FALSE
Buried L / Å	0.65	7.11	7.20	0.74	0.57	FALSE
OV_3 / Å^3	0.73	3.45	3.68	2.34	2.00	FALSE
$HV_{\uparrow}$	0.73	13.89	13.50	6.36	3.30	FALSE
L / Å	0.76	7.86	7.96	1.21	1.00	FALSE
OV_7 / Å^3	0.85	19.67	19.50	2.41	2.68	FALSE
OV_6 / Å^3	0.88	19.75	19.88	2.37	2.72	FALSE
E_{LUMO} / kJ mol^{-1}	1.00	16.1	16.1	55.7	43.6	FALSE

Table S5. Details of the ten reactant-intermediate pairs with lowest ΔE_{ML} selected from the two-step screening procedure, including the **Reaction Energy** (ΔE_{DFT}, target value), **Activation Energy** ($\Delta E_{\mathrm{DFT}}^{\ddagger}$), and the energy difference between the 6-coordinated intermediates with hydride located at equatorial position compared to hydride located at axial position, denoted as **Isomerization Energy**. For most structures of the intermediates, the isomer with hydride at the axial position is more energetically favorable. For structures of **L1**-**L10** see Table 1.

Reactant-intermediate pairs	**Reaction Energy ΔE_{DFT} (kJ mol^{-1})**	**Activation Energy $\Delta E_{\mathrm{DFT}}^{\ddagger}$ (kJ mol^{-1})**	**Isomerization Energy (kJ mol^{-1})**
L1	21.9	93.7	-47.3
L2	44.2	126.3	32.3
L3	19.8	89.0	-66.1
L4	36.2	114.9	0.7
L5	25.4	99.0	-37.4
L6	41.3	119.2	-22.3
L7	27.1	120.0	-43.1
L8	31.5	105.3	-33.5
L9	38.2	110.1	-36.9
L10	46.7	106.1	-17.2

References

1. C. Bannwarth, S. Ehlert and S. Grimme, GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions, *J. Chem. Theory Comput.*, 2019, **15**, 1652-1671.
2. J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615-6620.
3. F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
4. T. Lu and F. Chen, Multiwfn: a multifunctional wavefunction analyzer, *J. Comput. Chem.*, 2012, **33**, 580-592.
5. H. Schmider and A. Becke, Chemical content of the kinetic energy density, *J. Mol. Struct.: THEOCHEM*, 2000, **527**, 51-61.
6. R. Bader, *Atoms in molecules: a quantum theory* Oxford University Press, USA, 1994.
7. R. J. B. Chérif F. Matta, *The quantum theory of atoms in molecules: from solid state to DNA and drug design*, Wiley‐VCH Verlag GmbH & Co. KGaA, 2007.
8. A. P. Bartók, R. Kondor and G. Csányi, On representing chemical environments, *Phys. Rev. B*, 2013, **87**, 184115.
9. R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023.
10. H. Wickham, R. François, L. Henry and K. Müller, dplyr: A Grammar of Data Manipulation.*Journal*, 2022.
11. H. Wickham, *Stringr: Simple, consistent wrappers for common string operations*, 2023.
12. M. Gordon, S. Gragg and P. Konings, *HtmlTable: Advanced tables for markdown/html*, 2017.
13. G. R. Warnes, B. Bolker, T. Lumley and M. G. R. Warnes, Package ‘gtools’, *R Package version*, 2015, **3**.
14. K. Ushey, J. Allaire and Y. Tang, *reticulate: Interface to ’Python’*, 2020.
15. L. T. Kjell Jorner, *kjelljorner/morfeus: v0.7.2* Zenodo, 2022.