Supporting information: Beyond scaling relations for the description of catalytic materials

Mie Andersen,[*,†] Sergey V. Levchenko,[‡] Matthias Scheffler,[‡] and Karsten Reuter[†]

†*Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany*

‡*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin-Dahlem, Germany*

E-mail: mie.andersen@ch.tum.de

S1 Additional DFT computational details

For the smaller data set considering additional facets, we used a four-layered (3×3) supercell for the (111) facet, an eight-layered (2×3) supercell for the (110) facet, and a five-layered (3×3) supercell for the (100) facet. The structures of all facets are shown in Fig. S1 along with an overview of the considered adsorption sites. All DFT calculations were carried out as periodic slab calculations employing a vacuum region of 16 Å perpendicular to the surface and a dipole correction.[1] Calculations involving Ni were carried out spin-polarized. For the AB alloys, the periodicity of the crystal structure restricts the supercell sizes that can be constructed to $(1 \times N)$ with even N. For the AgPd alloy we calculated the adsorption energies of O and C at the bridge2-s site for N equal to two and four. Since the adsorption energies for the two cell sizes differed by only around 10 meV, we used the (1×2) supercell for all calculations. For all calculations the Brillouin zone was sampled with a (4×4) ((4×6) for AB alloys) $\boldsymbol{k}$-point grid and cutoffs of 500 eV and 5000 eV were used for the orbitals and the charge density, respectively.

The adsorption energy calculations were carried out by placing the adsorbate at the considered adsorption site in an initial geometry where the bonding atom was located about 2 Å from every metal atom within the site ensemble. For top (bridge) sites the fixing of one (two) adsorbate-metal bond distances leaves two (one) degrees of freedom to be defined. These were fixed by maximizing the distance between the bonding atom and the metal atoms neighboring the site ensemble, which ensured that the vector between the bonding atom and the geometrical center of the site ensemble was perpendicular to the local surface plane. For CO and CH the non-bonding atom (O or H) was placed at a distance of about 1.1 Å from the bonding atom at the position that maximized the distances to the neighboring surface atoms, which ensured that the molecular axis was perpendicular to the local surface plane. Since the the OH molecule often adsorbs in a tilted position, different positions of the non-bonding atom (H) were tested, corresponding to different tilt directions. Only the most stable tilt direction at a given site was added to the data set.

During the relaxation the bottom nine metal layers ((211) facet), two layers ((111) facet), four layers ((110) facet) or three layers ((100) facet) were fixed in their bulk-truncated positions, while the top layers and the adsorbate were relaxed until the maximum force on each atom fell below 0.03 eV/Å. Calculations in which the adsorbate moved to another site during the relaxation were discarded. An adsorbate was considered to have moved to a lower coordinated site if any adsorbate-metal bond within the site ensemble deviated by more than 10% from the others (a threshold of 20% was used for sites at alloy surfaces where the adsorbate coordinates to different metal atoms). Correspondingly, an adsorbate was considered to have moved to a higher coordinated site if any adsorbate-metal bond to the metal atoms neighboring the site ensemble had come within 10% of the adsorbate-metal bonds within the site ensemble.

For the larger data set considering the (211) facet all adsorption sites that correspond to local minima on the potential energy surface were included. These sites were identified in the following way. For CO and OH all adsorption sites defined in Fig. S1(a) were tried, while for the other adsorbates only selected sites were tried based on the known preferred adsorption sites from our previous work.[2] For C, CH and H the tested sites were the four-fold-*s* site, all three-fold coordinated sites and the bridge2-*s* site, while for O all three-fold coordinated sites and the bridge2-*s* site were tested. For the alloys all adsorption sites not equivalent by symmetry within each of the considered general site types were tested. An overview of the number of stable adsorption sites found for each material/site/adsorbate combination can be found in Table S1. For the smaller data set considering additional facets only one adsorption site was included for each adsorbate (see Table S2). For CH at the (110) facet no single adsorption site was found to be stable for all metals, and we therefore considered two different sites. For Au, Ir, Ru and Pt this was the threefold site, while for Cu, Ag, Pd, Ni and Rh this was the longbridge site. All adsorption energies were calculated as formation energies with respect to CH_4, H_2 and H_2O in the gas phase as is typically done for CO methanation reaction intermediates. The adsorption energies can be found in the file "data.txt" and the

relaxed coordinates can be found in the file "structures.tar".

The DFT calculations employed ultrasoft pseudopotentials taken from the Quantum ESPRESSO pseudopotential library. These were generated using the "atomic" code by A. Dal Corso in 2012 (v.5.0.2 svn rev. 9415). In order to test the accuracy of the employed pseudopotentials, we repeated some of the adsorption energy calculations using the full-potential, all-electron simulation package FHI-aims[3] and default tight settings for the basis set and integration grids. Since the BEEF-vdW functional is not available in FHI-aims, we instead based the comparison on the RPBE functional. All adsorption energies were evaluated using metal lattice constants optimized with Quantum ESPRESSO and BEEF-vdW and by relaxing the structures until the maximum force fell below 0.05 eV/Å. No significant differences in the final geometries were observed. The adsorption energies are compared in Table S3. It is seen that the RPBE Quantum ESPRESSO and FHI-aims calculations largely agree within about 10–30 meV, with the exception of the O adsorption energy on Pt, for which we find a larger difference of about 100 meV.

Comparing also the Quantum ESPRESSO results using the BEEF-vdW and RPBE functionals, it is seen that the two functionals deviate by up to about 100 meV. This can therefore be taken as a rough estimate for the spread in adsorption energy predictions of current state-of-the-art generalized-gradient-approximation DFT functionals. This is an important information for data-driven learning approaches, since it sets the limit for the predictive accuracy that it makes sense to pursue.

All primary features included in this study were obtained either from the literature (see Table S4), from DFT calculations of the bulk and surface structures of each metal (see Table S5), from simply counting atoms or bonds at a given site (see Table S6), or from DFT calculations specific for every material/site combination considered. The file "data.txt" contains, in addition to the adsorption energies, also all of the primary features calculated for the corresponding material/site combinations. The primary features calculated for the additional 32 AB alloys (those for which adsorption energies were not explicitly calculated

by DFT) can be found in the file "primary-features-AB-alloys.txt". For some of the primary features certain choices had to be made in order to arrive at a consistent definition of that feature. For the moments of the d-band we observed that especially the higher order moments were very sensitive to the number of empty bands included in the integral over the projected density of states (PDOS). We therefore consistently integrated up to the energy above the Fermi level where the PDOS had fallen below a threshold of 0.01 $\text{Å}^{-3}\text{eV}^{-1}$. The PDOS calculations were carried out using the smearing-free tetrahedron method, an energy spacing of 0.01 eV, and a $\boldsymbol{k}$-point grid of (14×14) ((14×21) for AB alloys). For the primary features involving the density of states at the Fermi level a smearing of 0.1 eV was used and the PDOS was averaged over the interval $\pm$ 0.1 eV around the Fermi level.

S2 Additional SISSO computational details

For constructing the $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$ feature spaces we made use of the set of algebraic/functional operators given in eq. 1.

$$\hat{\mathbf{H}}^{(m)} \equiv \{I, +, -, \times, /, \log, \exp, \exp-, ^{-1}, ^{2}, ^{3}, ^{6}, \sqrt{}, \sqrt[3]{}, |-|, \sin, \cos, \text{scd}\}, \tag{1}$$

where scd is the standard Cauchy distribution defined as:

$$\text{scd}(x) = \frac{1}{\pi(1+x^2)}. \tag{2}$$

The superscript m indicates that when applying $\hat{\mathbf{H}}^{(m)}$ to primary features ϕ_1 and ϕ_2 a dimensional analysis is performed, which ensures that only physically meaningful combinations are retained (e.g. only primary features with the same unit are added or subtracted).

As mentioned in the main text, the sparsifying ℓ_0 constraint is applied to a smaller feature subspace selected by a screening procedure (sure independence screening (SIS)), where the size of the subspace is equal to a user-defined SIS value times the dimension of the

descriptor. The SIS value is not an ordinary hyperparameter and its optimization through a validation data set is not straightforward. Ideally, one would want to search the entire feature space for the optimal descriptor. However, this is not computationally tractable since the computational cost of the sparsifying ℓ_0 constraint grows exponential with the size of the searched feature space. Instead, the SIS value should be chosen as large as computationally possible. It should be noted, though, that it is possible to get a better result (i.e. a lower training error) for a lower SIS value, since the used feature subspace for a lower SIS value is not necessarily a subspace of the used feature subspace for a higher SIS value. However, in general one would expect that there is a higher probability of finding a good descriptor if searching a larger subspace, which is also confirmed by the fact that the training errors in Fig. S4 tend to decrease for larger SIS values. In the present work we tested a number of SIS values up till the largest value that remained computationally tractable at each considered dimension. Since the validation data set (alloys or alloys and additional facets) is rather different from the training data set (pure fcc(211) metals) the best descriptor (i.e. the descriptor with the lowest RMSE on the validation data set) was not always found for the highest SIS value, as can be observed in Fig. S4.

The single-atom and AB alloys data were divided into validation and test data sets in the following way. For each adsorbate/alloy combination (six adsorbates times eight SA and AB alloys), we randomly partitioned the data (the different adsorption sites considered) into 50% validation and 50% test data. In case of an uneven number of data points for a given adsorbate/alloy combination, we alternately assigned the additional data point to validation and test, respectively. This procedure resulted in exactly 270 data points in each data set with roughly the same proportion of data from each adsorbate/alloy combination.

For the identification of a descriptor with a good compromise between accuracy for alloys and facets, we included 50% of the facets data in the validation data. The facets data set was divided into validation and test data sets in a similar way, i.e. for each adsorbate/facet combination (six adsorbates times three facets), we randomly partitioned the data (the nine

different metals considered) into 50% validation and 50% test data. This procedure resulted in exactly 81 data points in each data set with roughly the same proportion of data from each adsorbate/facet combination.

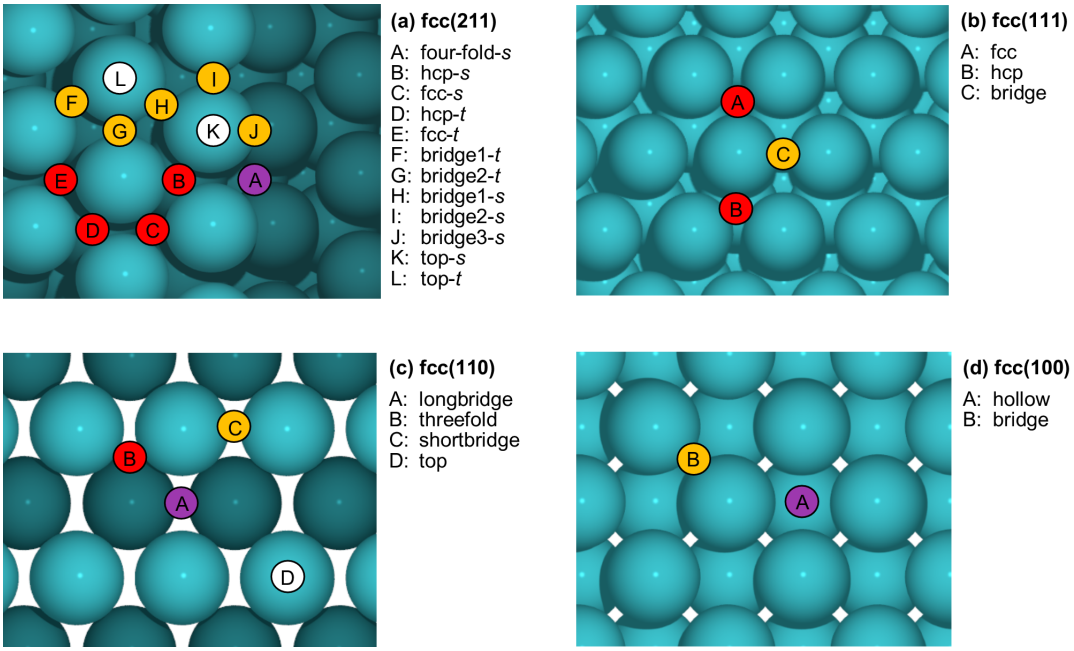


Figure S1: Overview of nomenclature used for the considered adsorption sites on (a) the (211), (b) the (111) facet, (c) the (110) facet and (d) the (100) facet.

Table S1: Overview of the number of stable adsorption sites identified for the larger data sets on the (211) facet.

Material class	Site class	C	O	H	CH	CO	OH	Total
Metals	Top	-	-	-	-	18	17	35
	Bridge	3	9	9	8	34	42	105
	Fcc and hcp	32	35	33	33	25	22	180
	Four-fold	9	-	3	9	3	-	24
	Total	44	44	45	50	80	81	**344**
Single-atom alloys	Top	-	-	-	-	16	14	30
	Bridge	3	7	8	7	29	43	97
	Fcc and hcp	26	28	25	25	20	14	138
	Four-fold	7	-	2	6	1	-	16
	Total	36	35	35	38	66	71	**281**
AB alloys	Top	-	-	-	-	16	10	26
	Bridge	3	7	8	7	23	52	100
	Fcc and hcp	22	30	17	29	6	9	113
	Four-fold	8	-	2	8	2	-	20
	Total	33	37	27	44	47	71	**259**

Table S2: Overview of considered adsorption sites for each adsorbate for the smaller data set considering additional facets.

	C	O	H	CH	CO	OH
(111)	hcp	fcc	fcc	hcp	hcp	bridge
(110)	longbridge	threefold	shortbridge	threefold/longbridge	top	shortbridge
(100)	hollow	hollow	hollow	hollow	bridge	bridge

Table S3: Comparison of adsorption energies (in eV) for O and H adsorption at the bridge2-*s* site calculated using various codes and functionals.

Code	Functional	Rh/O	Rh/H	Pt/O	Pt/H	Ag/O	Ag/H
Quantum ESPRESSO	BEEF-vdW	0.265	-0.351	1.130	-0.386	2.338	0.431
Quantum ESPRESSO	RPBE	0.375	-0.418	1.232	-0.465	2.463	0.353
FHI-aims	RPBE	0.343	-0.426	1.136	-0.454	2.440	0.331

Table S4: Primary features obtained from the literature. Pauling electronegativity (PE), ionization potential (IP in eV) and electron affinity (EA in eV) of the metal atom, radius of the *d*-orbitals in the bulk metal (r_d in Å) and coupling matrix element between the adsorbate states and the metal *d*-states squared (V_{ad}^2 relative to Cu).

	Cu	Pd	Pt	Rh	Ru	Ag	Au	Ir	Ni
PE[a]	1.90	2.20	2.28	2.28	2.20	1.93	2.54	2.20	1.91
IP[b]	7.726	8.337	8.959	7.459	7.361	7.576	9.226	9.121	7.640
EA[b]	1.236	0.562	2.125	1.143	1.046	1.304	2.309	1.564	1.157
r_d[c]	0.67	0.94	1.04	0.99	1.05	0.89	1.01	1.08	0.71
V_{ad}^2[d]	1	2.78	3.90	3.32	3.87	2.26	3.35	4.45	1.16

[a] From Ref.[4]
[b] From Ref.[5]
[c] From Ref.[6]
[d] From Ref.[7]

Table S5: Primary features obtained from DFT calculations of the bulk structure and fcc(211) surface of each material. Fcc nearest neighbor distance ($\mathrm{bulk}_{\mathrm{nnd}}$ in Å) and work function (W in eV).

	Cu	Pd	Pt	Rh	Ru	Ag	Au	Ir	Ni
$\mathrm{bulk}_{\mathrm{nnd}}$	2.603	2.819	2.838	2.734	2.732	2.985	2.990	2.753	2.502
W	4.719	5.210	5.750	5.078	4.889	4.427	5.280	5.489	4.945

Table S6: Primary features obtained by counting atoms or bonds at a given site. Number of metal atoms in site ensemble ($\mathrm{site}_{\mathrm{no}}$) and average coordination number of metal atoms in site ensemble (CN).

	four-fold-*s*	hcp-*s*	fcc-*s*	hcp-*t*	fcc-*t*	bridge1-*t*	bridge2-*t*	bridge1-*s*	bridge2-*s*	bridge3-*s*	top-*s*	top-*t*
$\mathrm{site}_{\mathrm{no}}$	4	3	3	3	3	2	2	2	2	2	1	1
CN	8.5	7.667	8.333	9.333	9.667	9.5	9	8	7	8.5	7	9

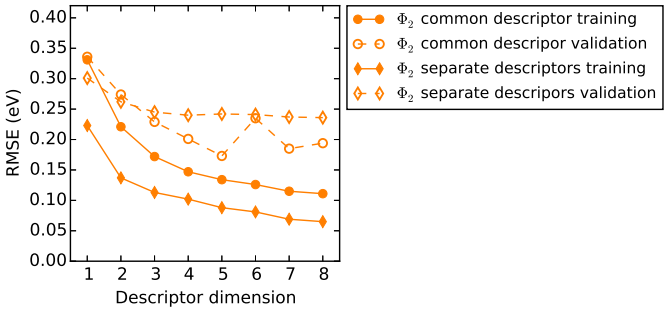
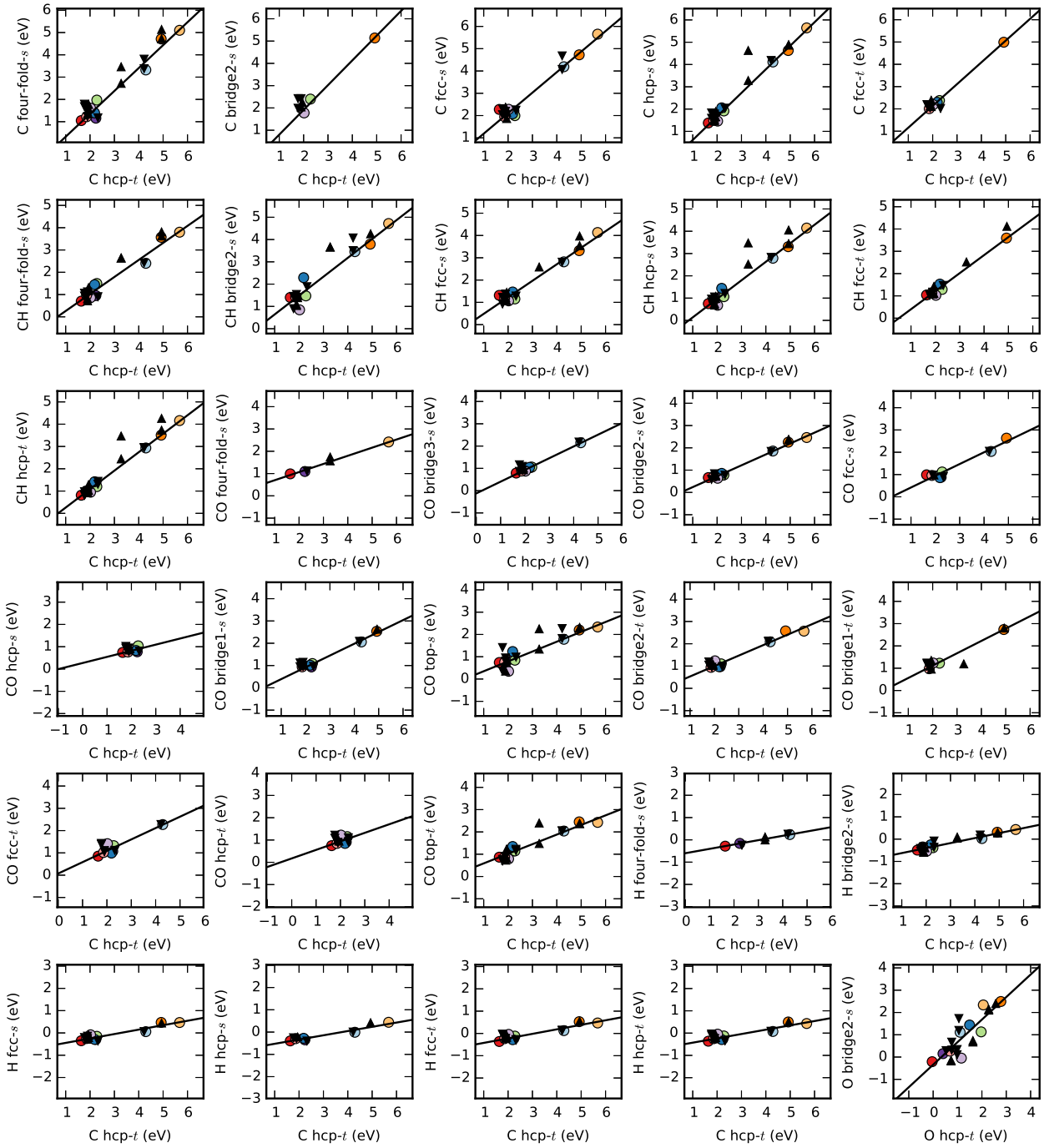


Figure S2: RMSE for the identified descriptors of $\mathbf{\Phi}_2$ for multi-task learning of a common descriptor for all adsorbates (dots) and for the learning of separate descriptors for each adsorbate (rhombi). The pure metals data set is used for training and 50% of the alloys data are used for validation. The SIS values used are 2D: 20000, 3D: 1000, 4D: 100, 5D: 50, 6D: 10, 7D: 10, and 8D: 8.

Table S7: Largest deviations between calculated and predicted adsorption energies for the best SISSO descriptor (8D, $\mathbf{\Phi}_3$, SIS=9) and for the combined validation and test data sets containing all single-atom and AB alloys. Energies are in eV. Note that for the alloys there are several (up to four) in-equivalent sites of each site type labeled 1,2,3,4.

Alloy	Site	Adsorbate	Calculated ads. energy	Predicted ads. energy	Deviation
Au@Ni	top-*s*-1	OH	0.98	1.59	0.61
AgPd	four-fold-*s*-1	C	2.72	3.25	0.54
AgAu	hcp-*t*-1	C	4.94	5.47	0.53
Ag@Cu	bridge2-*s*-1	CH	4.07	4.59	0.51
AgAu	hcp-*s*-2	C	4.89	5.39	0.50
Pd@Ir	bridge2-*s*-1	C	2.42	2.92	0.50
Ag@Cu	fcc-*s*-1	C	4.68	5.10	0.42
AgAu	four-fold-*s*-2	C	4.72	5.14	0.42
Ag@Cu	four-fold-*s*-1	C	3.81	4.21	0.41
Pd@Ir	bridge2-*s*-1	O	0.34	0.74	0.40


C four-fold-s (eV)
C bridge2-s (eV)
C fcc-s (eV)
C hcp-s (eV)
C fcc-t (eV)
CH four-fold-s (eV)
CH bridge2-s (eV)
CH fcc-s (eV)
CH hcp-s (eV)
CH fcc-t (eV)
CH hcp-t (eV)
CO four-fold-s (eV)
CO bridge3-s (eV)
CO bridge2-s (eV)
CO fcc-s (eV)
CO hcp-s (eV)
CO bridge1-s (eV)
CO top-s (eV)
CO bridge2-t (eV)
CO bridge1-t (eV)
CO fcc-t (eV)
CO hcp-t (eV)
CO top-t (eV)
H four-fold-s (eV)
H bridge2-s (eV)
H fcc-s (eV)
H hcp-s (eV)
H fcc-t (eV)
H hcp-t (eV)
O bridge2-s (eV)
C hcp-t (eV)
O hcp-t (eV)

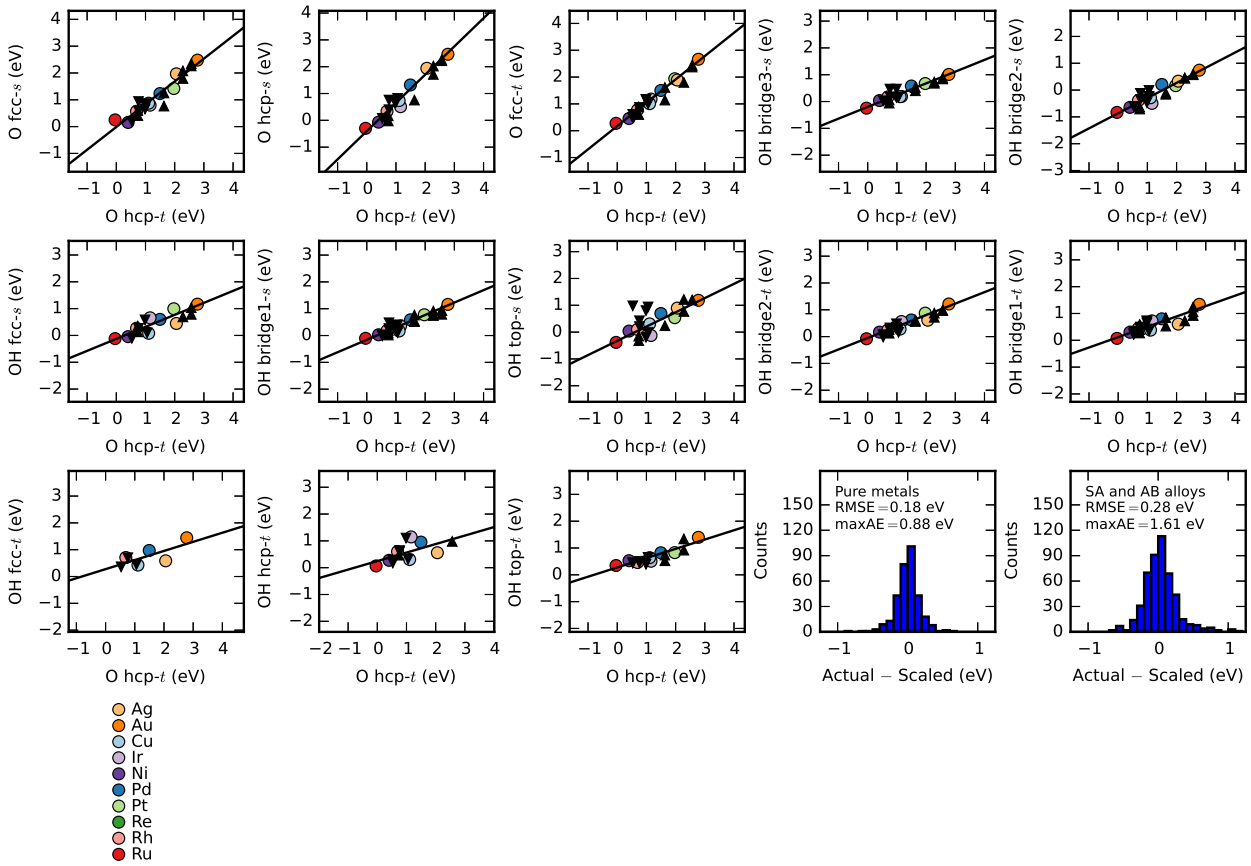


Figure S3: Scaling relations for all adsorbate/site combinations. The black lines are fits to the DFT adsorption energies at the pure metals (colored dots). Explicitly calculated DFT adsorption energies for the SA (AB) alloys are shown as black triangles pointing down (up). The last two plots show the error distribution for the least-squares fitting to the metals as well as for the validation on the alloys. Finally, the color coding used for the metals is shown.

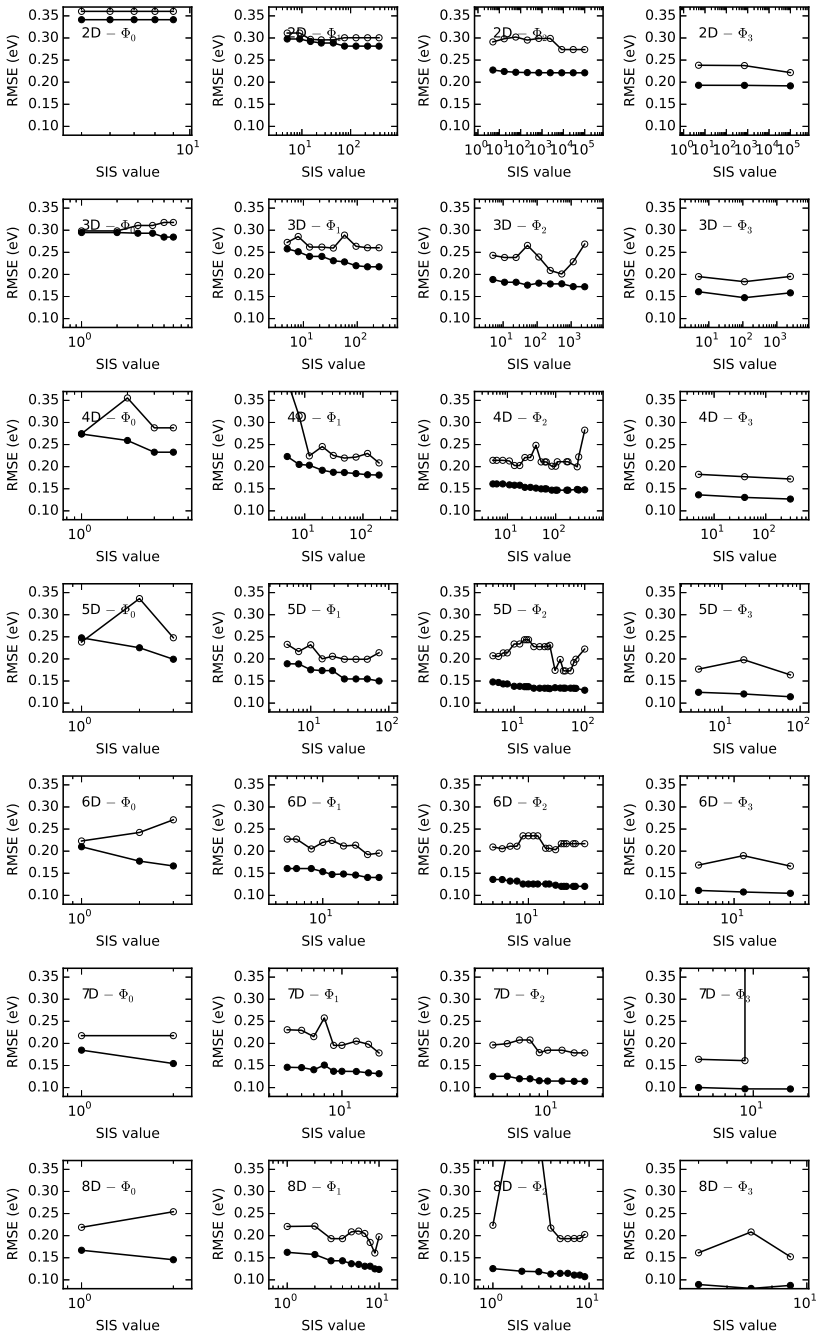


Figure S4: RMSE for the identified 2D-8D descriptors of $\mathbf{\Phi}_0$, $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$ and $\mathbf{\Phi}_3$ as a function of the SIS value. Training (validation) errors are shown with filled (open) symbols. For the smaller feature spaces ($\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ up to 4D) the range of possible SIS values is limited by the constraint that the size of the selected feature subspace cannot be larger than the size of the full feature space. For the larger feature spaces the range of possible SIS values is limited by the computational cost of the sparsifying ℓ_0 constraint, which grows exponential with the feature subspace size.

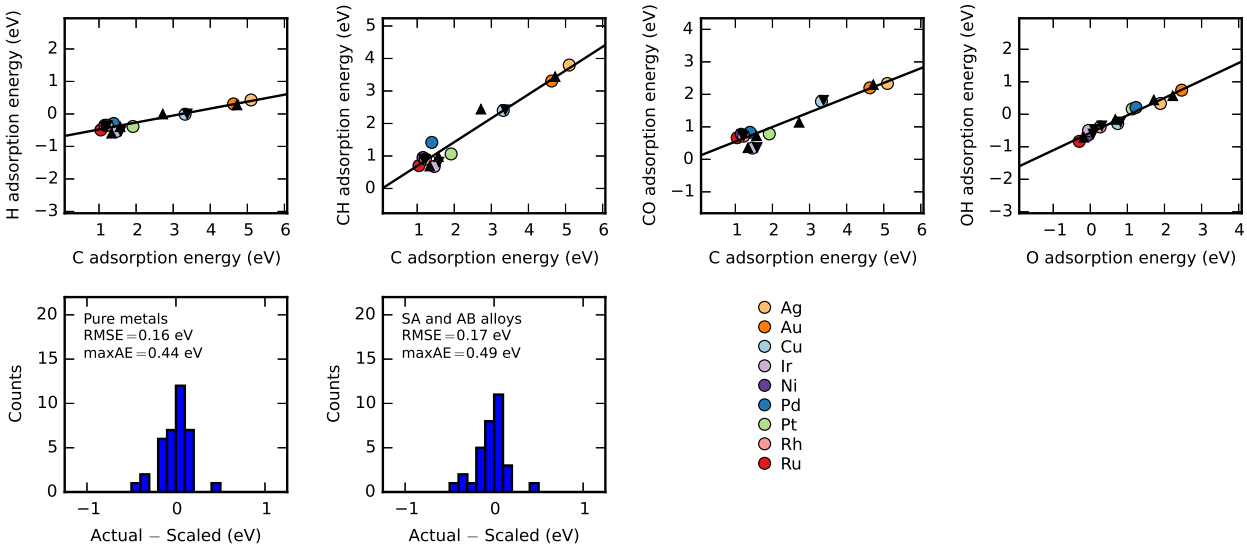


Figure S5: Scaling relations when considering only the most stable adsorption site of each adsorbate at each metal or alloy. The black lines are fits to the DFT adsorption energies at the pure metals (colored dots). Explicitly calculated DFT adsorption energies for the SA (AB) alloys are shown as black triangles pointing down (up). The last two plots show the error distribution for the least-squares fitting to the metals as well as for the validation on the alloys. Finally, the color coding used for the metals is shown.

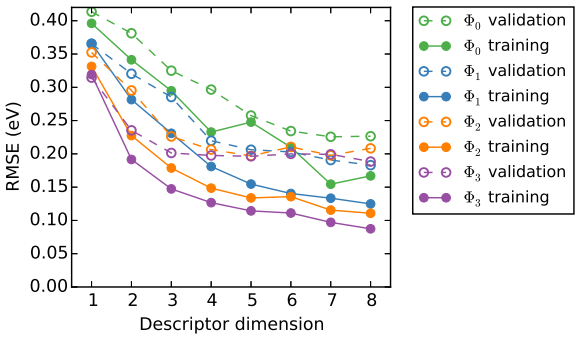


Figure S6: RMSE for the descriptors identified using exclusively the pure metals data set for training and 50% of the alloys and facets data sets for validation.

Table S8: Descriptors identified using exclusively the pure metals data set for training and 50% of the alloys data for validation. For each $\mathbf{\Phi}_n$ (n $\in$ {0,1,2,3}) and for each dimension d^m (m $\in$ {1,2,3,4,5,6,7,8}) we give the descriptor with the lowest RMSE on the validation data set out of the SIS values tested (see Fig. S4). The fitting coefficients for each adsorbate can be found in the file "Descriptor-coefficients-training-metals-validation-alloys.txt".

n	d^m	Descriptor	SIS value
0	d_1^1	(ε_d)	–
0	d_1^2	(ε_d)	9
	d_2^2	(W)	
0	d_1^3	(ε_d)	2
	d_2^3	(W)	
	d_3^3	(f_d)	
0	d_1^4	(ε_d)	1
	d_2^4	(W)	
	d_3^4	(f_d)	
	d_4^4	(CN)	
0	d_1^5	(ε_d)	1
	d_2^5	(W)	
	d_3^5	(f_d)	
	d_4^5	(CN)	
	d_5^5	$(\text{site}_{\text{no}})$	
0	d_1^6	(ε_d)	1
	d_2^6	(W)	
	d_3^6	(f_d)	
	d_4^6	(CN)	
	d_5^6	$(\text{site}_{\text{no}})$	
	d_6^6	(PE)	
0	d_1^7	(ε_d)	1

	d_2^7	(W)	
	d_3^7	(f_d)	
	d_4^7	(CN)	
	d_5^7	$(\mathrm{site_{no}})$	
	d_6^7	(PE)	
	d_7^7	(f_{sp})	
0	d_1^8	(ε_d)	1
	d_2^8	(W)	
	d_3^8	(f_d)	
	d_4^8	(CN)	
	d_5^8	$(\mathrm{site_{no}})$	
	d_6^8	(PE)	
	d_7^8	(f_{sp})	
	d_8^8	(S_d)	
1	d_1^1	$(\varepsilon_d \cdot f_d)$	–
1	d_1^2	$(\varepsilon_d \cdot f_d)$	25
	d_2^2	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	
1	d_1^3	$(\varepsilon_d \cdot f_d)$	36
	d_2^3	$\sin(\mathrm{DOS}_d)$	
	d_3^3	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	
1	d_1^4	$(\mathrm{PE} \cdot \varepsilon_d)$	189
	d_2^4	$(W \cdot \mathrm{site_{nnd}})$	
	d_3^4	$(f_d/\mathrm{site_{no}})$	
	d_4^4	$(W_d \cdot \mathrm{DOS}_{sp})$	
1	d_1^5	$(W_d + \varepsilon_d)$	27
	d_2^5	$(W)^3$	
	d_3^5	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	

	d_4^5	$(\text{site}_{\text{no}}/\text{PE})$	
	d_5^5	$(\text{site}_{\text{no}} \cdot f_{sp})$	
1	d_1^6	$(W_d + \varepsilon_d)$	24
	d_2^6	$(W)^3$	
	d_3^6	$(\text{DOS}_{sp}/\text{PE})$	
	d_4^6	$(\text{site}_{\text{no}}/\text{PE})$	
	d_5^6	$(\text{site}_{\text{no}})^6$	
	d_6^6	$(\text{site}_{\text{no}} \cdot f_{sp})$	
1	d_1^7	(ε_d)	16
	d_2^7	(ε_d/W)	
	d_3^7	$(\text{DOS}_{sp}/\text{IP})$	
	d_4^7	$\text{cbrt}(W)$	
	d_5^7	$(\text{site}_{\text{no}} \cdot \text{bulk}_{\text{nnd}})$	
	d_6^7	$(\text{site}_{\text{nnd}}/\text{site}_{\text{no}})$	
	d_7^7	$(V_{\text{ad}}^2/\text{site}_{\text{no}})$	
1	d_1^8	$(\varepsilon_d/\text{bulk}_{\text{nnd}})$	9
	d_2^8	$(W \cdot \text{bulk}_{\text{nnd}})$	
	d_3^8	$(\text{IP} \cdot \text{CN})$	
	d_4^8	$(W)^3$	
	d_5^8	$(\text{site}_{\text{no}} \cdot W)$	
	d_6^8	$\text{sqrt}(\text{site}_{\text{no}})$	
	d_7^8	$(V_{\text{ad}}^2 \cdot \text{DOS}_{sp})$	
	d_8^8	$(\text{EA}/V_{\text{ad}}^2)$	
2	d_1^1	$(\cos(r_d) \cdot (\text{PE} \cdot \varepsilon_d))$	–
2	d_1^2	$((f_d)^3/\text{cbrt}(\text{DOS}_d))$	100000
	d_2^2	$(\text{sqrt}(\text{DOS}_{sp})/(\text{PE} \cdot W))$	
2	d_1^3	$(\text{scd}(\text{DOS}_d) \cdot (\varepsilon_d \cdot f_d))$	529

	d_2^3	$((\mathrm{PE} \cdot f_d) \cdot \log(\mathrm{DOS}_{sp}))$	
	d_3^3	$((f_d/\mathrm{site_{no}})/(\mathrm{IP} - \varepsilon_d))$	
2	d_1^4	$((\mathrm{PE} \cdot \varepsilon_d)/\mathrm{sqrt}(r_d))$	280
	d_2^4	$(\exp(\mathrm{DOS}_{sp})/(W \cdot \mathrm{bulk_{nnd}}))$	
	d_3^4	$(\mathrm{scd}(\mathrm{site_{no}})/(r_d \cdot f_{sp}))$	
	d_4^4	$((\mathrm{IP} \cdot r_d)/(\mathrm{site_{no}} + \mathrm{CN}))$	
2	d_1^5	$((\mathrm{IP} \cdot f_d)/\mathrm{cbrt}(\mathrm{DOS}_d))$	53
	d_2^5	$((W \cdot \mathrm{site_{nnd}})/\exp(\mathrm{DOS}_{sp}))$	
	d_3^5	$(\mathrm{cbrt}(K_d)/(\mathrm{site_{no}} \cdot f_{sp}))$	
	d_4^5	$(\sin(\mathrm{site_{no}})/(W_d \cdot \varepsilon_d))$	
	d_5^5	$((W_d \cdot \varepsilon_d)/(\mathrm{CN})^6)$	
2	d_1^6	$((\mathrm{PE} \cdot \mathrm{bulk_{nnd}})/(r_d/\varepsilon_d))$	17
	d_2^6	$((\mathrm{PE} \cdot f_d)/\mathrm{cbrt}(\mathrm{DOS}_{sp}))$	
	d_3^6	$(\exp(-\mathrm{site_{no}})/(W_d \cdot \varepsilon_d))$	
	d_4^6	$((W_d \cdot f_{sp})/\exp(\mathrm{CN}))$	
	d_5^6	$((f_{sp})^6 \cdot (V_{\mathrm{ad}}^2 \cdot \mathrm{site_{no}}))$	
	d_6^6	$((r_d \cdot \mathrm{bulk_{nnd}})/(\mathrm{PE})^2)$	
2	d_1^7	$((\mathrm{PE} \cdot \mathrm{bulk_{nnd}})/(r_d/\varepsilon_d))$	16
	d_2^7	$((\mathrm{PE} \cdot f_d)/\mathrm{cbrt}(\mathrm{DOS}_{sp}))$	
	d_3^7	$(\exp(-\mathrm{DOS}_{sp})/(\mathrm{site_{no}} \cdot W))$	
	d_4^7	$((V_{\mathrm{ad}}^2 \cdot f_{sp})/\exp(\mathrm{CN}))$	
	d_5^7	$((r_d \cdot \mathrm{site_{no}})/\exp(\mathrm{CN}))$	
	d_6^7	$(\exp(V_{\mathrm{ad}}^2)/\sin(\mathrm{site_{no}}))$	
	d_7^7	$(\mathrm{abs}(\mathrm{DOS}_d - \mathrm{DOS}_{sp})/\exp(\varepsilon_d))$	
2	d_1^8	$(\mathrm{scd}(r_d) \cdot (\mathrm{PE} \cdot \varepsilon_d))$	6
	d_2^8	$((\mathrm{PE} \cdot f_d)/\mathrm{sqrt}(\mathrm{DOS}_{sp}))$	
	d_3^8	$((\mathrm{CN}/f_{sp})/(\mathrm{site_{no}})^2)$	

	d_4^8	$((V_{\mathrm{ad}}^2/\mathrm{CN}))^6$	
	d_5^8	$((W_d \cdot f_{sp})/\sin(\mathrm{site_{no}}))$	
	d_6^8	$((W_d/r_d)/\sin(\mathrm{site_{no}}))$	
	d_7^8	$\mathrm{abs}(\mathrm{PE} - (\mathrm{DOS}_{sp}/\mathrm{DOS}_d))$	
	d_8^8	$(\log(W_d)/\cos(\mathrm{CN}))$	
3	d_1^1	$(((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\varepsilon_d)) \cdot ((\varepsilon_d + W)/(f_d \cdot \mathrm{site_{nnd}})))$	–
3	d_1^2	$(((V_{\mathrm{ad}}^2/\mathrm{PE}) - \sin(\mathrm{site_{no}})) + ((\mathrm{IP}/\varepsilon_d) \cdot \sin(\varepsilon_d)))$	100000
	d_2^2	$(((\mathrm{PE} \cdot f_d) - (f_{sp}/V_{\mathrm{ad}}^2)) \cdot (\log(\mathrm{DOS}_{sp}) \cdot (\varepsilon_d - W)))$	
3	d_1^3	$(((\mathrm{EA} + \mathrm{IP}) \cdot (r_d/\varepsilon_d)) \cdot (\sin(\varepsilon_d)/(\mathrm{site_{nnd}})^2))$	112
	d_2^3	$((\log(\mathrm{DOS}_{sp}) \cdot (r_d \cdot f_d)) \cdot ((f_d + f_{sp}) - (f_d/\mathrm{PE})))$	
	d_3^3	$((\sin(\mathrm{site_{no}}) \cdot (\mathrm{site_{no}}/V_{\mathrm{ad}}^2))/((f_{sp})^2 \cdot (\mathrm{CN})^6))$	
3	d_1^4	$(((\mathrm{IP} \cdot W) \cdot \mathrm{sqrt}(W)) \cdot (\sin(\varepsilon_d)/(\varepsilon_d \cdot f_d)))$	300
	d_2^4	$((\log(\mathrm{DOS}_{sp}) \cdot (r_d \cdot f_d)) \cdot ((f_d + f_{sp}) - (f_d/\mathrm{PE})))$	
	d_3^4	$((\sin(\mathrm{site_{no}}) \cdot (\mathrm{site_{no}}/V_{\mathrm{ad}}^2))/((f_{sp})^2 \cdot (\mathrm{CN})^6))$	
	d_4^4	$\mathrm{abs}((\cos(r_d) \cdot (V_{\mathrm{ad}}^2 - \mathrm{PE})) - (\sin(W_d) + \cos(\mathrm{site_{no}})))$	
3	d_1^5	$((\sin(\varepsilon_d) + \mathrm{scd}(f_{sp})) \cdot ((\mathrm{IP} \cdot W)/(\varepsilon_d \cdot f_d)))$	75
	d_2^5	$((\log(\mathrm{DOS}_{sp}) \cdot (r_d \cdot f_d)) \cdot ((f_d + f_{sp}) - (f_d/\mathrm{PE})))$	
	d_3^5	$((\sin(\mathrm{site_{no}})/(\mathrm{IP})^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^5	$((\sin(\mathrm{CN})/\cos(\mathrm{site_{no}})) - (\log(W_d)/\cos(\mathrm{CN})))$	
	d_5^5	$(((f_{sp})^6 \cdot (\mathrm{CN}/f_d)) \cdot \mathrm{abs}((\mathrm{PE} \cdot \mathrm{site_{no}}) - (\mathrm{CN}/V_{\mathrm{ad}}^2)))$	
3	d_1^6	$(((\mathrm{EA})^2 + (\mathrm{IP} \cdot W)) \cdot (\sin(\varepsilon_d)/(\varepsilon_d \cdot f_d)))$	30
	d_2^6	$(((\mathrm{PE} \cdot W) + (W_d - \varepsilon_d)) \cdot (\log(\mathrm{DOS}_{sp}) \cdot (f_d \cdot \mathrm{bulk_{nnd}})))$	
	d_3^6	$((\mathrm{sqrt}(\mathrm{DOS}_d)/(\mathrm{DOS}_d + \mathrm{DOS}_{sp})) \cdot \mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^6	$((\sin(\mathrm{site_{no}})/(\mathrm{IP})^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_5^6	$(\mathrm{abs}(V_{\mathrm{ad}}^2 - (\mathrm{CN}/\mathrm{site_{no}})) - (\log(W_d)/\cos(\mathrm{CN})))$	
	d_6^6	$\mathrm{abs}((\mathrm{scd}(\varepsilon_d) \cdot \sin(\mathrm{CN})) - ((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\mathrm{DOS}_{sp})))$	
3	d_1^7	$(((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\varepsilon_d)) \cdot ((\varepsilon_d + W)/(f_d \cdot \mathrm{site_{nnd}})))$	9

	d_2^7	$((\mathrm{PE} + \mathrm{cbrt}(V_{\mathrm{ad}}^2)) \cdot ((f_d + f_{sp})/\mathrm{cbrt}(\mathrm{DOS}_{sp})))$	
	d_3^7	$((\sin(\mathrm{site_{no}})/(W)^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^7	$((\cos(\mathrm{CN}) \cdot (\mathrm{site_{no}} \cdot f_{sp})) \cdot \mathrm{abs}((V_{\mathrm{ad}}^2 \cdot \mathrm{site_{no}}) - (\mathrm{CN} - \mathrm{site_{no}})))$	
	d_5^7	$((\sin(\mathrm{EA})/\cos(\mathrm{site_{no}}))/((W_d/\varepsilon_d) - \sin(\mathrm{site_{no}})))$	
	d_6^7	$((\sin(\mathrm{PE}) + \cos(f_d))/\sin((\mathrm{site_{no}} \cdot \mathrm{CN})))$	
	d_7^7	$\mathrm{abs}((\cos(\mathrm{PE}) + \sin(\mathrm{EA})) - (\sin(\mathrm{IP}) + \cos(\mathrm{site_{no}})))$	
3	d_1^8	$(((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\varepsilon_d)) \cdot ((\varepsilon_d + W)/(f_d \cdot \mathrm{bulk_{nnd}})))$	9
	d_2^8	$((\mathrm{PE} + \mathrm{cbrt}(V_{\mathrm{ad}}^2)) \cdot ((f_d + f_{sp})/\mathrm{cbrt}(\mathrm{DOS}_{sp})))$	
	d_3^8	$((\sin(\mathrm{site_{no}})/(W)^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^8	$((\cos(\mathrm{CN}) \cdot (\mathrm{site_{no}} \cdot f_{sp})) \cdot \mathrm{abs}((V_{\mathrm{ad}}^2 \cdot \mathrm{site_{no}}) - (\mathrm{CN} - \mathrm{site_{no}})))$	
	d_5^8	$((\sin(\mathrm{EA})/\cos(\mathrm{site_{no}}))/((W_d/\varepsilon_d) - \sin(\mathrm{site_{no}})))$	
	d_6^8	$((\sin(\mathrm{PE}) + \cos(f_d))/\sin((\mathrm{site_{no}} \cdot \mathrm{CN})))$	
	d_7^8	$\mathrm{abs}((\sin(\mathrm{EA}))^6 - (\sin(\mathrm{IP}) + \cos(\mathrm{site_{no}})))$	
	d_8^8	$(\sin((\mathrm{CN})^2) \cdot ((r_d \cdot V_{\mathrm{ad}}^2) \cdot (\varepsilon_d + W)))$	

Table S9: Descriptors identified using exclusively the pure metals data set for training and 50% of the alloys and facets data sets for validation. For each $\mathbf{\Phi}_n$ (n $\in$ {0,1,2,3}) and for each dimension d^m (m $\in$ {1,2,3,4,5,6,7,8}) we give the descriptor with the lowest RMSE on the validation data set out of the SIS values tested (see Fig. S4). The fitting coefficients for each adsorbate can be found in the file "Descriptor-coefficients-training-metals-validation-alloys-facets.txt".

n	d^m	Descriptor	SIS value
0	d_1^1	(ε_d)	–
0	d_1^2	(ε_d)	9
	d_2^2	(W)	
0	d_1^3	(ε_d)	2
	d_2^3	(W)	
	d_3^3	(f_d)	
0	d_1^4	(ε_d)	3
	d_2^4	(V_{ad}^2)	
	d_3^4	(DOS_{sp})	
	d_4^4	(PE)	
0	d_1^5	(ε_d)	1
	d_2^5	(W)	
	d_3^5	(f_d)	
	d_4^5	(CN)	
	d_5^5	$(\mathrm{site_{no}})$	
0	d_1^6	(ε_d)	1
	d_2^6	(W)	
	d_3^6	(f_d)	
	d_4^6	(CN)	
	d_5^6	$(\mathrm{site_{no}})$	
	d_6^6	(PE)	

0	d_1^7	(ε_d)	2
	d_2^7	(DOS_d)	
	d_3^7	(W)	
	d_4^7	(f_d)	
	d_5^7	$(\mathrm{site_{no}})$	
	d_6^7	(DOS_{sp})	
	d_7^7	(PE)	
0	d_1^8	(ε_d)	1
	d_2^8	(W)	
	d_3^8	(f_d)	
	d_4^8	(CN)	
	d_5^8	$(\mathrm{site_{no}})$	
	d_6^8	(PE)	
	d_7^8	(f_{sp})	
	d_8^8	(S_d)	
1	d_1^1	$(\varepsilon_d \cdot f_d)$	–
1	d_1^2	$\log(\mathrm{DOS}_d)$	75
	d_2^2	$(f_d \cdot \mathrm{site_{nnd}})$	
1	d_1^3	$(\varepsilon_d \cdot f_d)$	36
	d_2^3	$\sin(\mathrm{DOS}_d)$	
	d_3^3	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	
1	d_1^4	$(\mathrm{PE} \cdot \varepsilon_d)$	189
	d_2^4	$(W \cdot \mathrm{site_{nnd}})$	
	d_3^4	$(f_d/\mathrm{site_{no}})$	
	d_4^4	$(W_d \cdot \mathrm{DOS}_{sp})$	
1	d_1^5	$(W_d + \varepsilon_d)$	27

	d_2^5	$(W)^3$	
	d_3^5	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	
	d_4^5	$(\mathrm{site_{no}}/\mathrm{PE})$	
	d_5^5	$(\mathrm{site_{no}} \cdot f_{sp})$	
1	d_1^6	$(W_d + \varepsilon_d)$	30
	d_2^6	$(W)^3$	
	d_3^6	$(\mathrm{DOS}_{sp}/\mathrm{PE})$	
	d_4^6	$(\mathrm{site_{no}}/\mathrm{PE})$	
	d_5^6	$(\mathrm{site_{no}} \cdot f_{sp})$	
	d_6^6	$(\mathrm{DOS}_{sp}/\mathrm{site_{no}})$	
1	d_1^7	$(W_d + \varepsilon_d)$	14
	d_2^7	$(W \cdot \mathrm{site_{nnd}})$	
	d_3^7	$(\mathrm{site_{no}}/\mathrm{PE})$	
	d_4^7	$(\mathrm{site_{no}})^2$	
	d_5^7	$(f_{sp} \cdot \mathrm{CN})$	
	d_6^7	(DOS_{sp}/f_d)	
	d_7^7	$(\mathrm{bulk_{nnd}}/\mathrm{PE})$	
1	d_1^8	$(\varepsilon_d/\mathrm{bulk_{nnd}})$	9
	d_2^8	$(W \cdot \mathrm{bulk_{nnd}})$	
	d_3^8	$(\mathrm{IP} \cdot \mathrm{CN})$	
	d_4^8	$(W)^3$	
	d_5^8	$(\mathrm{site_{no}} \cdot W)$	
	d_6^8	$\mathrm{sqrt}(\mathrm{site_{no}})$	
	d_7^8	$(V_{\mathrm{ad}}^2 \cdot \mathrm{DOS}_{sp})$	
	d_8^8	$(\mathrm{EA}/V_{\mathrm{ad}}^2)$	
2	d_1^1	$(\cos(r_d) \cdot (\mathrm{PE} \cdot \varepsilon_d))$	–
2	d_1^2	$(\cos(r_d) \cdot (\mathrm{PE} \cdot \varepsilon_d))$	5

	d_2^2	$(\mathrm{cbrt}(\mathrm{DOS}_{sp})/(\mathrm{IP}\cdot\mathrm{site}_{\mathrm{nnd}}))$	
2	d_1^3	$(\mathrm{scd}(\mathrm{DOS}_d)\cdot(\varepsilon_d\cdot f_d))$	529
	d_2^3	$((\mathrm{PE}\cdot f_d)\cdot\log(\mathrm{DOS}_{sp}))$	
	d_3^3	$((f_d/\mathrm{site}_{\mathrm{no}})/(\mathrm{IP}-\varepsilon_d))$	
2	d_1^4	$((\mathrm{PE}\cdot\varepsilon_d)/\mathrm{sqrt}(r_d))$	280
	d_2^4	$(\exp(\mathrm{DOS}_{sp})/(W\cdot\mathrm{bulk}_{\mathrm{nnd}}))$	
	d_3^4	$(\mathrm{scd}(\mathrm{site}_{\mathrm{no}})/(r_d\cdot f_{sp}))$	
	d_4^4	$((\mathrm{IP}\cdot r_d)/(\mathrm{site}_{\mathrm{no}}+\mathrm{CN}))$	
2	d_1^5	$((\mathrm{IP}\cdot f_d)/\mathrm{cbrt}(\mathrm{DOS}_d))$	53
	d_2^5	$((W\cdot\mathrm{site}_{\mathrm{nnd}})/\exp(\mathrm{DOS}_{sp}))$	
	d_3^5	$(\mathrm{cbrt}(K_d)/(\mathrm{site}_{\mathrm{no}}\cdot f_{sp}))$	
	d_4^5	$(\sin(\mathrm{site}_{\mathrm{no}})/(W_d\cdot\varepsilon_d))$	
	d_5^5	$((W_d\cdot\varepsilon_d)/(\mathrm{CN})^6)$	
2	d_1^6	$((\mathrm{PE}\cdot\varepsilon_d)/\exp(r_d))$	6
	d_2^6	$(\mathrm{cbrt}(\mathrm{DOS}_{sp})/(\mathrm{IP}\cdot\mathrm{site}_{\mathrm{nnd}}))$	
	d_3^6	$((\mathrm{CN}/f_{sp})/(\mathrm{site}_{\mathrm{no}})^2)$	
	d_4^6	$(\sin(\mathrm{site}_{\mathrm{no}})/(\mathrm{EA}\cdot r_d))$	
	d_5^6	$(\exp(V_{\mathrm{ad}}^2)\cdot\exp(-\mathrm{CN}))$	
	d_6^6	$\mathrm{abs}((W_d/K_d)-(\mathrm{site}_{\mathrm{no}}/\mathrm{CN}))$	
2	d_1^7	$((\mathrm{PE}\cdot\mathrm{bulk}_{\mathrm{nnd}})/(r_d/\varepsilon_d))$	9
	d_2^7	$((\mathrm{PE}\cdot f_d)/\mathrm{cbrt}(\mathrm{DOS}_{sp}))$	
	d_3^7	$(\exp(-\mathrm{DOS}_{sp})/(\mathrm{site}_{\mathrm{no}}\cdot W))$	
	d_4^7	$((V_{\mathrm{ad}}^2\cdot f_{sp})/\exp(\mathrm{CN}))$	
	d_5^7	$((r_d\cdot\mathrm{site}_{\mathrm{no}})/\exp(\mathrm{CN}))$	
	d_6^7	$((V_{\mathrm{ad}}^2)^6/\sin(\mathrm{site}_{\mathrm{no}}))$	
	d_7^7	$(\mathrm{abs}(\mathrm{DOS}_d-\mathrm{DOS}_{sp})/\exp(\varepsilon_d))$	
2	d_1^8	$(\mathrm{scd}(r_d)\cdot(\mathrm{PE}\cdot\varepsilon_d))$	8

	d_2^8	$(\mathrm{cbrt}(\mathrm{DOS}_{sp})/(\mathrm{IP}\cdot\mathrm{site}_{\mathrm{nnd}}))$	
	d_3^8	$(\sin(\mathrm{IP})\cdot\sin(\mathrm{site}_{\mathrm{no}}))$	
	d_4^8	$((V_{\mathrm{ad}}^2/\mathrm{CN}))^6$	
	d_5^8	$((\mathrm{site}_{\mathrm{no}}/f_{sp})/\cos(\mathrm{site}_{\mathrm{no}}))$	
	d_6^8	$(\exp(\mathrm{CN})/\cos(\mathrm{CN}))$	
	d_7^8	$\mathrm{abs}((W_d/K_d)-(\mathrm{site}_{\mathrm{no}}/\mathrm{CN}))$	
	d_8^8	$((W_d/S_d)/\mathrm{abs}(\mathrm{DOS}_d-\mathrm{DOS}_{sp}))$	
3	d_1^1	$(((\mathrm{site}_{\mathrm{no}}/\mathrm{CN})-\sin(\varepsilon_d))\cdot((\varepsilon_d+W)/(f_d\cdot\mathrm{site}_{\mathrm{nnd}})))$	–
3	d_1^2	$(((V_{\mathrm{ad}}^2/\mathrm{PE})-\sin(\mathrm{site}_{\mathrm{no}}))+((\mathrm{IP}/\varepsilon_d)\cdot\sin(\varepsilon_d)))$	100000
	d_2^2	$(((\mathrm{PE}\cdot f_d)-(f_{sp}/V_{\mathrm{ad}}^2))\cdot(\log(\mathrm{DOS}_{sp})\cdot(\varepsilon_d-W)))$	
3	d_1^3	$(((\mathrm{EA}+\mathrm{IP})\cdot(r_d/\varepsilon_d))\cdot(\sin(\varepsilon_d)/(\mathrm{site}_{\mathrm{nnd}})^2))$	112
	d_2^3	$((\log(\mathrm{DOS}_{sp})\cdot(r_d\cdot f_d))\cdot((f_d+f_{sp})-(f_d/\mathrm{PE})))$	
	d_3^3	$((\sin(\mathrm{site}_{\mathrm{no}})\cdot(\mathrm{site}_{\mathrm{no}}/V_{\mathrm{ad}}^2))/((f_{sp})^2\cdot(\mathrm{CN})^6))$	
3	d_1^4	$(((\mathrm{IP}\cdot W)\cdot\mathrm{sqrt}(W))\cdot(\sin(\varepsilon_d)/(\varepsilon_d\cdot f_d)))$	300
	d_2^4	$((\log(\mathrm{DOS}_{sp})\cdot(r_d\cdot f_d))\cdot((f_d+f_{sp})-(f_d/\mathrm{PE})))$	
	d_3^4	$((\sin(\mathrm{site}_{\mathrm{no}})\cdot(\mathrm{site}_{\mathrm{no}}/V_{\mathrm{ad}}^2))/((f_{sp})^2\cdot(\mathrm{CN})^6))$	
	d_4^4	$\mathrm{abs}((\cos(r_d)\cdot(V_{\mathrm{ad}}^2-\mathrm{PE}))-(\sin(W_d)+\cos(\mathrm{site}_{\mathrm{no}})))$	
3	d_1^5	$((\sin(\varepsilon_d)+\mathrm{scd}(f_{sp}))\cdot((\mathrm{IP}\cdot W)/(\varepsilon_d\cdot f_d)))$	75
	d_2^5	$((\log(\mathrm{DOS}_{sp})\cdot(r_d\cdot f_d))\cdot((f_d+f_{sp})-(f_d/\mathrm{PE})))$	
	d_3^5	$((\sin(\mathrm{site}_{\mathrm{no}})/(\mathrm{IP})^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site}_{\mathrm{no}})-(f_{sp}/f_d)))$	
	d_4^5	$((\sin(\mathrm{CN})/\cos(\mathrm{site}_{\mathrm{no}}))-(\log(W_d)/\cos(\mathrm{CN})))$	
	d_5^5	$(((f_{sp})^6\cdot(\mathrm{CN}/f_d))\cdot\mathrm{abs}((\mathrm{PE}\cdot\mathrm{site}_{\mathrm{no}})-(\mathrm{CN}/V_{\mathrm{ad}}^2)))$	
3	d_1^6	$(((\mathrm{site}_{\mathrm{no}}/\mathrm{CN})-\sin(\varepsilon_d))\cdot((\varepsilon_d+W)/(f_d\cdot\mathrm{site}_{\mathrm{nnd}})))$	5
	d_2^6	$((\mathrm{PE}+\mathrm{cbrt}(V_{\mathrm{ad}}^2))\cdot((f_d+f_{sp})/\mathrm{cbrt}(\mathrm{DOS}_{sp})))$	
	d_3^6	$((\sin(\mathrm{site}_{\mathrm{no}})/(W)^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site}_{\mathrm{no}})-(f_{sp}/f_d)))$	
	d_4^6	$(((\mathrm{site}_{\mathrm{no}})^3\cdot\mathrm{abs}(\mathrm{PE}-V_{\mathrm{ad}}^2))\cdot(\cos(\mathrm{CN})/(\mathrm{CN})^6))$	
	d_5^6	$((\sin(\mathrm{EA})/\cos(\mathrm{site}_{\mathrm{no}}))/((W_d/\varepsilon_d)-\sin(\mathrm{site}_{\mathrm{no}})))$	

	d_6^6	$((\sin(\mathrm{PE}) + \cos(f_d))/\sin((\mathrm{site_{no}} \cdot \mathrm{CN})))$	
3	d_1^7	$(((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\varepsilon_d)) \cdot ((\varepsilon_d + W)/(f_d \cdot \mathrm{site_{nnd}})))$	9
	d_2^7	$((\mathrm{PE} + \mathrm{cbrt}(V_{\mathrm{ad}}^2)) \cdot ((f_d + f_{sp})/\mathrm{cbrt}(\mathrm{DOS}_{sp})))$	
	d_3^7	$((\sin(\mathrm{site_{no}})/(W)^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^7	$((\cos(\mathrm{CN}) \cdot (\mathrm{site_{no}} \cdot f_{sp})) \cdot \mathrm{abs}((V_{\mathrm{ad}}^2 \cdot \mathrm{site_{no}}) - (\mathrm{CN} - \mathrm{site_{no}})))$	
	d_5^7	$((\sin(\mathrm{EA})/\cos(\mathrm{site_{no}}))/((W_d/\varepsilon_d) - \sin(\mathrm{site_{no}})))$	
	d_6^7	$((\sin(\mathrm{PE}) + \cos(f_d))/\sin((\mathrm{site_{no}} \cdot \mathrm{CN})))$	
	d_7^7	$\mathrm{abs}((\cos(\mathrm{PE}) + \sin(\mathrm{EA})) - (\sin(\mathrm{IP}) + \cos(\mathrm{site_{no}})))$	
3	d_1^8	$(((\mathrm{site_{no}}/\mathrm{CN}) - \sin(\varepsilon_d)) \cdot ((\varepsilon_d + W)/(f_d \cdot \mathrm{bulk_{nnd}})))$	9
	d_2^8	$((\mathrm{PE} + \mathrm{cbrt}(V_{\mathrm{ad}}^2)) \cdot ((f_d + f_{sp})/\mathrm{cbrt}(\mathrm{DOS}_{sp})))$	
	d_3^8	$((\sin(\mathrm{site_{no}})/(W)^6)/\mathrm{abs}(\mathrm{scd}(\mathrm{site_{no}}) - (f_{sp}/f_d)))$	
	d_4^8	$((\cos(\mathrm{CN}) \cdot (\mathrm{site_{no}} \cdot f_{sp})) \cdot \mathrm{abs}((V_{\mathrm{ad}}^2 \cdot \mathrm{site_{no}}) - (\mathrm{CN} - \mathrm{site_{no}})))$	
	d_5^8	$((\sin(\mathrm{EA})/\cos(\mathrm{site_{no}}))/((W_d/\varepsilon_d) - \sin(\mathrm{site_{no}})))$	
	d_6^8	$((\sin(\mathrm{PE}) + \cos(f_d))/\sin((\mathrm{site_{no}} \cdot \mathrm{CN})))$	
	d_7^8	$\mathrm{abs}((\sin(\mathrm{EA}))^6 - (\sin(\mathrm{IP}) + \cos(\mathrm{site_{no}})))$	
	d_8^8	$(\sin((\mathrm{CN})^2) \cdot ((r_d \cdot V_{\mathrm{ad}}^2) \cdot (\varepsilon_d + W)))$	

Table S10: Descriptor identified using using the pooled metals and alloys data sets excluding the 23 DFT-calculated alloy data points from Fig. 2(a) and (b) for training. The fitting coefficients for each adsorbate can be found in the file "Descriptor-coefficients-training-pooled-minus-exclusions.txt".

n	d^m	Descriptor	SIS value
3	d_1^8	$((\sin(f_d)/\sin(\text{bulk}_{\text{nnd}})) - ((\text{IP}/\varepsilon_d) + \sin(\text{site}_{\text{no}})))$	9
	d_2^8	$(((r_d \cdot f_{sp})/\text{scd}(\text{PE})) - (\log(\text{DOS}_{sp}) \cdot (f_d \cdot \text{bulk}_{\text{nnd}})))$	
	d_3^8	$\text{abs}(((\text{PE} \cdot \text{site}_{\text{no}})/(W_d - W)) - (\log(f_{sp}) \cdot (\text{CN}/W_d)))$	
	d_4^8	$(\sin((\text{PE} \cdot \text{site}_{\text{no}}))/((\text{site}_{\text{no}} - \text{CN}))^6)$	
	d_5^8	$((\cos(\text{site}_{\text{no}}) \cdot (W/W_d)) - (\exp(V_{\text{ad}}^2) \cdot \text{scd}(\text{site}_{\text{no}})))$	
	d_6^8	$((\exp(-\text{EA})/\exp(f_d)) \cdot \cos((\text{site}_{\text{no}} \cdot \text{CN})))$	
	d_7^8	$\cos((\exp(W)/(\varepsilon_d \cdot f_d)))$	
	d_8^8	$(((\text{IP} \cdot \text{PE}) \cdot (\text{CN}/\varepsilon_d))/\text{cbrt}(\text{abs}(\text{DOS}_d - \text{DOS}_{sp})))$	

Table S11: Descriptor identified using using the pooled metals and alloys data sets for training. The fitting coefficients for each adsorbate can be found in the file "Descriptor-coefficients-training-pooled.txt".

n	d^m	Descriptor	SIS value
3	d_1^8	$((\sin(f_d)/\sin(\text{bulk}_{\text{nnd}})) - ((\text{IP}/\varepsilon_d) + \sin(\text{site}_{\text{no}})))$	9
	d_2^8	$(((r_d \cdot f_{sp})/\text{scd}(\text{PE})) - (\log(\text{DOS}_{sp}) \cdot (f_d \cdot \text{bulk}_{\text{nnd}})))$	
	d_3^8	$(\sin((\text{site}_{\text{nnd}}/\text{EA})) \cdot ((\text{CN})^3 \cdot (\text{site}_{\text{no}}/W)))$	
	d_4^8	$(\sin((\text{PE} \cdot \text{site}_{\text{no}})) \cdot (\exp(\text{site}_{\text{no}})/(\text{CN})^6))$	
	d_5^8	$((\exp(V_{\text{ad}}^2) \cdot \text{scd}(\text{site}_{\text{no}})) - (\cos(\text{site}_{\text{no}}) \cdot (\text{bulk}_{\text{nnd}}/r_d)))$	
	d_6^8	$(\cos((\text{site}_{\text{no}} \cdot \text{CN}))/(\exp(f_d) \cdot (W_d \cdot \varepsilon_d)))$	
	d_7^8	$(((\text{IP} \cdot \text{PE})/(\text{EA} \cdot W_d)) \cdot ((f_{sp}/f_d) - \text{scd}(\text{PE})))$	
	d_8^8	$\text{abs}((\cos(\text{DOS}_d) \cdot (W/\text{EA})) - (\text{abs}(\text{PE} - V_{\text{ad}}^2)/\cos(\text{CN})))$	

References

(1) Bengtsson, L. Dipole Correction for Surface Supercell Calculations. *Phys. Rev. B* **1999**, *59*, 12301–12304.

(2) Andersen, M.; Plaisance, C. P.; Reuter, K. Assessment of Mean-Field Microkinetic Models for CO Methanation on Stepped Metal Surfaces Using Accelerated Kinetic Monte Carlo. *J. Chem. Phys.* **2017**, *147*, 152705.

(3) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175 – 2196.

(4) Winter, M. WebElements. `https://www.webelements.com/`, 2018 (accessed May 25, 2018).

(5) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5*, 24131–24138.

(6) Harrison, W. A. *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*; Dover Publications, 1989.

(7) Ruban, A.; Hammer, B.; Stoltze, P.; Skriver, H.; Nørskov, J. Surface Electronic Structure and Reactivity of Transition and Noble Metals. *J. Mol. Catal. A: Chem.* **1997**, *115*, 421 – 429.