Supplementary Information: Interpretable Bayesian Optimization for Catalyst Discovery

Akhil S. Nair,[1,2] Lucas Foppa,[1] and Matthias Scheffler[1]

[1] The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany

[2] Institut für Chemie und Biochemie, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany

Table S1: List of features in the SAA dataset used in the study. The features are characterized based on whether they are related to the host, dopant single-atom, surface site, local coordination or adsorption.

| Type	Symbol	Unit	Description
Host			
	PE_h	-	host Pauling electronegativity
	IP_h	eV	host ionization potential
	EA_h	eV	host electron affinity
	r_{s-h}	Å	host s-orbital atomic radius
	r_{p-h}	Å	host p-orbital atomic radius
	r_{d-h}	Å	host d-orbital atomic radius
	r_{val-h}	Å	host valence radius
	bulk_{h-nnd}	Å	neighbor distance in host bulk
Single-atom			
	PE_{SA}	-	SA Pauling electronegativity
	IP_{SA}	eV	SA ionization potential
	EA_{SA}	eV	SA electron affinity
	r_{s-SA}	Å	SA s-orbital atomic radius
	r_{p-SA}	Å	SA p-orbital atomic radius
	r_{d-SA}	Å	SA d-orbital atomic radius
	r_{val-SA}	Å	SA valence radius
Surface site			
	PE_{site}	-	surface site PE[a]
	IP_{site}	eV	surface site IP[a]
	EA_{site}	eV	surface site EA[a]
	site_{no}	# atoms	atoms in the surface site
Local coordination			
	PE_{snn}	-	surface site and first neighbors PE[a]
	IP_{snn}	eV	surface site and first neighbors IP[a]
	EA_{snn}	eV	surface site and first neighbors EA[a]
	CN	# atoms	surface site coordination number
	gen-CN	# atoms	generalized CN
Adsorption			
	E_{ads}	eV	Adsorption energy of CO_2
	O-C-O_{ang}	degree	O-C-O bond angle

(a)   (b)   (c)
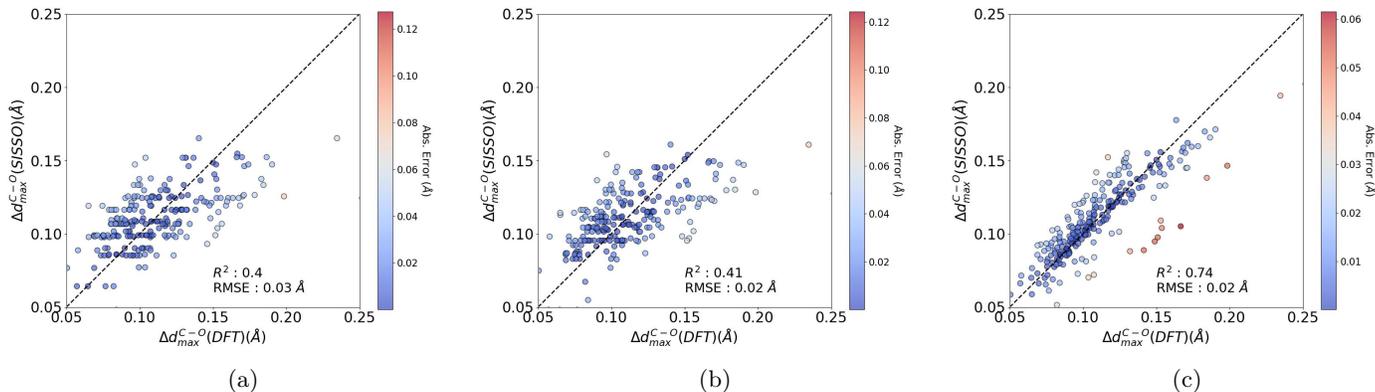
Fig. S1: Training performance of SISSO model (a) without using adsorption-related properties as features, (b) with the hi-SISSO approach adding the model and descriptors obtained from another SISSO model for adsorption energy prediction to the set of features and (c) including adsorption-related properties directly as features. The x- and y-axes of the parity plots show the true and predicted maximum C-O bond elongation ($\Delta d_{\max}^{\mathrm{C-O}}$) values, respectively.

Table S2 : Features selected by different feature selection methods for predicting $\Delta d_{\max}^{\mathrm{C-O}}$. PE denotes Pauling electronegativity, IP ionization potential, EA electron affinity, rs/rp/rd/rval represent s-, p-, d-orbital and valence radii respectively, CN coordination number, and subscripts indicate: h (host), sa (single-atom), site (surface site), snn (surface site with first neighbors in the coordination environment). The features related to host, single-atom, site, adsorption and local coordination are indicated by cyan, orange, magenta, green and violet colours, respectively.

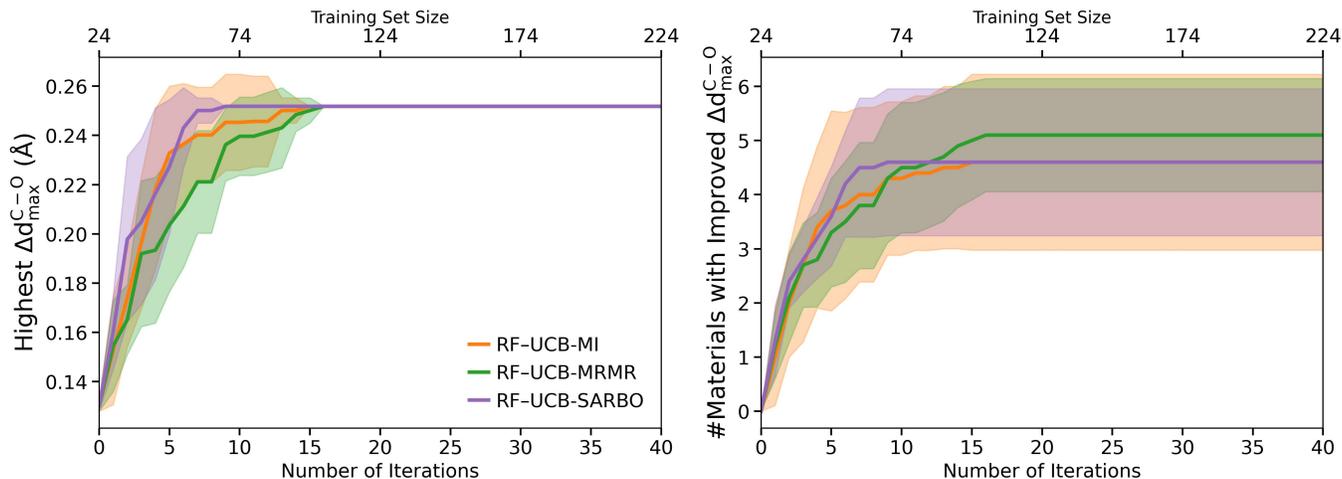| FS Method	Selected Features
SISSO	CN, OCO_{ang}, PE_h, PE_{sa}, gen_{CN}, rs_h, $site_{no}$
MI	OCO_{ang}, PE_{site}, $rval_{sa}$, EA_{site}, IP_{site}, rd_{sa}, rs_{sa}
RFE	E_{ads}, OCO_{ang}, EA_{sa}, rd_{sa}, IP_{snn}, EA_{snn}, gen_{CN}
MRMR	EA_{site}, EA_{sa}, PE_{sa}, PE_{site}, rd_{sa}, E_{ads}, OCO_{ang}

Fig. S2. Performance comparison between BO with different feature selection strategies using RF as the surrogate model and UCB as the acqusition function: (left) maximum $\Delta d_{\max}^{\mathrm{C-O}}$ evolution, (right) cumulative improvements in target property. Shaded regions indicate standard deviation across 10 independent trials.