

## Supplementary information for:

### Site-averaged kinetics for catalysts on amorphous supports: an importance learning algorithm

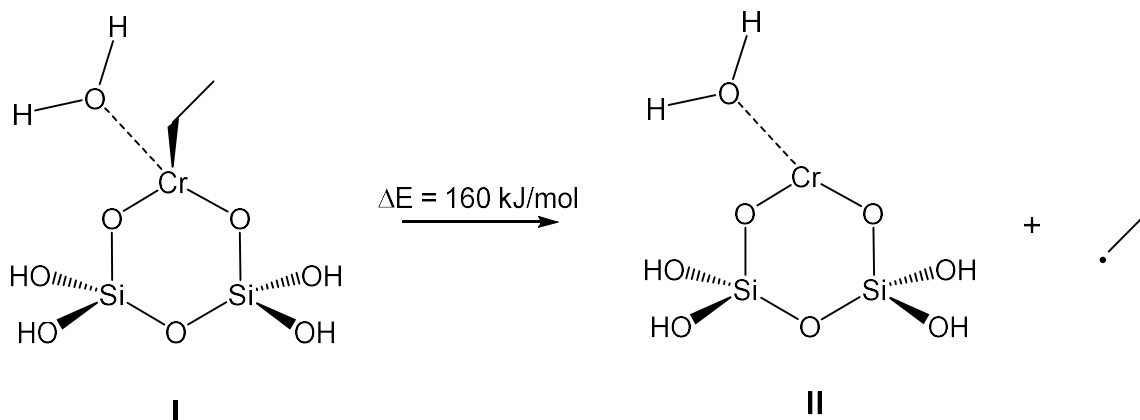
*Craig A. Vandervelden,<sup>a,‡</sup> Salman A. Khan,<sup>a,‡</sup> Susannah L. Scott,<sup>a,b</sup> Baron Peters<sup>c</sup>*

- Department of Chemical Engineering, University of California, Santa Barbara, California 93106-5080.
  - Department of Chemistry, University of California, Santa Barbara, California 93106-5080.
  - Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, Illinois 61801.
- ‡. Authors with equal contribution.

TABLE OF CONTENTS		Page
<b>Section S1</b>	Strength of M-A bond.	S1
<b>Scheme S1</b>	Reaction pathway to calculate Cr-C bond strength.	S2
<b>Table S1</b>	Spin contamination before (S2) and after annihilation of highest spin contaminant (S2A). Cartesian coordinates of optimized structures used to estimate M-A bond strength	S2 S3
<b>Section S2</b>	Derivation of apparent activation energy.	S4
<b>Section S3</b>	Propagation of kernel regression model uncertainty in estimating $\langle E_a \rangle_k$	S5
<b>Section S4</b>	Test set and training set statistics	S7
<b>Fig. S1</b>	Parity plot of kernel regression model trained on different initial pool sizes.	S7
<b>Fig. S2</b>	Kernel regression model residual distributions trained on different initial pool sizes.	S8
<b>Fig. S3</b>	Residual distribution for iterations 0 and 30 of the importance learning algorithm.	S9
<b>Section S5</b>	Sample size vs error for $\bar{E}_a$ estimates by reweighting	S9
<b>References</b>	Additional references	S11

#### S1. Strength of M-A bond

The M-A bond strength in the quenched-disordered lattice model was chosen to approximately match the Cr-C bond strength for an alkylchromium(III) site on SiO<sub>2</sub>. We started from a bis(silanolato)chromium(II) cluster model, which has been used in previous studies of Cr/SiO<sub>2</sub> catalysts.<sup>1,2</sup> Labile siloxane coordination was modeled by binding a water molecule. The M-A bond strength was calculated according to **Scheme S1** and density functional theory calculations, and was computed as  $\epsilon_{\text{M-A}} = 160$  kJ/mol. We chose a one-electron redox pathway (as opposed to a two-electron redox pathway) to ensure a strong bond energy for the chemisorption step. We stress that the bond strength was chosen only to ensure a realistic model. The model is not intended make accurate predictions for Cr/SiO<sub>2</sub> olefin polymerization catalysts.



**Scheme S1:** Reaction pathway to calculate Cr-C bond strength.

DFT calculations were carried out using Gaussian16.<sup>3</sup> All energies were calculated with the range-separated density functional,  $\omega$ B97X-D.<sup>4</sup> The def2-TZVP basis set was used for Cr<sup>5</sup> and TZVP was used for C, H, O, and Si atoms.<sup>6</sup> All minima have zero imaginary frequencies. The peripheral OH atoms of the cluster model were also held constrained to model the geometric constraints of an extended silica network. The peripheral atom constraints were found by optimizing the bare Cr(II) cluster. The same peripheral atom constraints were applied to structures **I** and **II**. Cartesian coordinates of the optimized clusters are tabulated below **Table S1**.

**Table S1:** Spin contamination before (S2) and after annihilation (S2A) of highest spin contaminant; energies in Hartrees

Species	Energy	S2	S2A
<b>I</b>	-2229.395021	6.0107	6.0000
<b>II</b>	-2308.613831	3.8158	3.7506
<b>C<sub>3</sub>H<sub>7</sub></b>	-157.7869992	0.7544	0.7500

Cartesian coordinates of optimized structures used to estimate **M-A** bond strength

**I**

O	1.528715	1.824430	0.017273
O	0.329040	3.677023	1.481622
O	-3.798037	2.422373	0.623959
O	-2.437772	0.962573	2.367903
H	1.471874	1.112840	0.626215
H	0.721382	3.163392	2.193211
H	-3.813726	2.949265	0.179675
H	-3.058776	0.288005	2.079504
Cr	-0.437283	4.457655	2.661415
O	0.530001	4.130430	1.122581
O	-2.025354	3.604722	2.371739
O	-1.123096	2.190320	0.310740
Si	-2.349940	2.320287	1.425008
Si	0.314150	2.956189	0.007436
O	0.963648	6.104979	2.702537
H	0.804577	7.049654	2.679910
H	1.377310	5.837656	1.865089
C	0.720489	2.007273	3.871771
C	0.399543	3.460992	4.187553
H	1.291691	4.017253	4.491492
H	-0.347290	3.537112	4.990328
H	1.457606	1.942161	3.067310
H	-0.170998	1.454563	3.568652
H	1.144256	1.492268	4.740030

**II**

O	1.528715	1.824430	-0.017273
O	0.329040	3.677023	-1.481622
O	-3.798037	2.422373	0.623959
O	-2.437772	0.962573	2.367903
H	1.471874	1.112840	0.626215
H	0.721382	3.163392	-2.193211
H	-3.813726	2.949265	-0.179675
H	-3.058776	0.288005	2.079504
Cr	-0.583821	4.714397	2.500341
O	0.579509	4.129526	1.136052
O	-2.065920	3.613509	2.402040
O	-1.121537	2.226881	0.331971
Si	-2.356280	2.335183	1.444016
Si	0.326075	2.970978	0.018794
H	1.148930	6.972860	2.668851
O	1.081087	6.020796	2.752306
H	1.518486	5.602420	1.984661

**C<sub>3</sub>H<sub>7</sub>**

C	-2.607098	4.751562	-3.928794
C	-3.579273	5.597963	-1.753988
C	-3.191053	4.390545	-2.611791
C	-2.379889	6.436179	-1.325945
H	-1.985557	4.055188	-4.476072
H	-4.086401	3.767304	-2.765517
H	-4.282718	6.220844	-2.315828
H	-4.116340	5.248218	-0.868301
H	-1.680277	5.839589	-0.734938
H	-1.836621	6.816586	-2.193793
H	-2.687355	7.290288	-0.719788
H	-2.492122	3.754728	-2.055856
H	-2.905619	5.665313	-4.429382

## S2. Derivation of apparent activation energy

In this section, a formula for the apparent activation energy of a site,  $E_a(\mathbf{x}_i)$ , is derived. The apparent activation energy for site  $i$  is given by

$$E_a(\mathbf{x}_i) = -\frac{d \ln \kappa}{d\beta} \quad (\text{S1})$$

From eqns. (2) – (5), the turnover frequency (TOF) of a site,  $\kappa_i$ , can be expressed as

$$\kappa = k_2 K(\mathbf{x}_i) c_A = \frac{k_B T}{h} \exp\left[-\frac{\Delta H^\ddagger - T\Delta S^\ddagger}{k_B T}\right] \exp\left[-\frac{\Delta H(\mathbf{x}_i) - T\Delta S}{k_B T}\right] c_A \quad (\text{S2})$$

Taking the natural logarithm of the eqn. (S2), grouping temperature dependent terms, and simplifying yields

$$\ln \kappa = \ln c_A - \ln h + \frac{\Delta S + \Delta S^\ddagger}{k_B} - \beta \Delta H(\mathbf{x}_i, \beta) - \beta \Delta H^\ddagger - \ln \beta \quad (\text{S3})$$

where  $\beta = 1/k_B T$ . From eqn. (6),  $\Delta H(\mathbf{x}_i)$  is temperature dependent through  $k_B T$ . Taking the derivative of eqn. (S3) gives

$$E_a(\mathbf{x}_i) = \frac{d}{d\beta} \beta \Delta H(\mathbf{x}_i, \beta) + \Delta H^\ddagger + k_B T \quad (\text{S4})$$

Inserting eqn. (6) into  $\Delta H(\mathbf{x}_i)$  to evaluate the derivative gives

$$\begin{aligned} \frac{d}{d\beta} \beta \Delta H(\mathbf{x}_i, \beta) &= \frac{d}{d\beta} [\beta V_{AM^*}(\mathbf{x}_i) - \beta V_{M^*}(\mathbf{x}_i) + 1] \\ &= V_{AM^*}(\mathbf{x}_i) - V_{M^*}(\mathbf{x}_i) \\ &= \Delta H(\mathbf{x}_i) + k_B T \end{aligned} \quad (\text{S5})$$

Thus,  $E_a(\mathbf{x}_i)$  can be written as

$$E_a(\mathbf{x}_i) = \Delta H(\mathbf{x}_i) + \Delta H^\ddagger + 2k_B T \quad (\text{S6})$$

### S3. Propagation of kernel regression model uncertainty in estimating $\langle E_a \rangle_k$

Site-averaged kinetics are estimated by importance sampling the activation energy distribution with  $E_a$  values obtained from the trained kernel regression model. Since errors in the kernel regression model propagate through the  $\langle E_a \rangle_k$  calculation (beyond sampling error and error from *ab initio* calculations), the kernel regression model contributes additional errors. Here, we show that the regression errors, even when unbiased, will systematically bias the  $\langle E_a \rangle_k$  estimate toward lower activation energy. We also show how this bias can be quantified and corrected to obtain  $\langle E_a \rangle_k$  estimates with only sampling and *ab initio* calculation errors. Let the distribution of kernel regression activation energies be  $\tilde{\rho}(\hat{E}_a)$ .  $\tilde{\rho}(\hat{E}_a)$  can be related to the  $E_a$  distribution,  $\tilde{\rho}(E_a)$ , by

$$\tilde{\rho}(\hat{E}_a) = \int dE_a \tilde{\rho}(E_a) P(\hat{E}_a | E_a). \quad (\text{S7})$$

Here,  $P(\hat{E}_a | E_a)$  is the distribution of the model-predicted activation barriers around the true activation barriers, and the integral is over the all possible  $E_a$  values. The site averaged activation energy from  $\tilde{\rho}(\hat{E}_a)$  is

$$\langle \hat{E}_a \rangle_k = \frac{\int d\hat{E}_a \hat{E}_a e^{-\beta \hat{E}_a} \tilde{\rho}(\hat{E}_a)}{\int d\hat{E}_a e^{-\beta \hat{E}_a} \tilde{\rho}(\hat{E}_a)}. \quad (\text{S8})$$

where  $\beta = 1/k_B T$ , and  $T$  is the operating temperature of the catalyst. Combining eqns (S7) and (S8) yields

$$\langle \hat{E}_a \rangle_k = \frac{\int \int d\hat{E}_a dE_a \hat{E}_a e^{-\beta \hat{E}_a} P(\hat{E}_a | E_a) \tilde{\rho}(E_a)}{\int \int d\hat{E}_a dE_a e^{-\beta \hat{E}_a} P(\hat{E}_a | E_a) \tilde{\rho}(E_a)}. \quad (\text{S9})$$

Assuming  $\tilde{\rho}(\hat{E}_a)$  is normally distributed around  $\tilde{\rho}(E_a)$  with a standard deviation of  $\sigma_g$  gives:

$$\tilde{\rho}(\hat{E}_a) = \int_{-\infty}^{+\infty} dE_a \tilde{\rho}(E_a) \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left[-\frac{(\hat{E}_a - E_a)^2}{2\sigma_g^2}\right]. \quad (\text{S10})$$

Combining eqns (S9) and (S10) and simplifying yields:

$$\langle \hat{E}_a \rangle_k = \frac{\int dE_a \tilde{\rho}(E_a) \int_{-\infty}^{+\infty} d\hat{E}_a \hat{E}_a \exp\left[\frac{(\hat{E}_a - E_a)^2}{2\sigma_g^2} - \beta \hat{E}_a\right]}{\int dE_a \tilde{\rho}(E_a) \int_{-\infty}^{+\infty} d\hat{E}_a \exp\left[\frac{(\hat{E}_a - E_a)^2}{2\sigma_g^2} - \beta \hat{E}_a\right]}. \quad (\text{S11})$$

The two integrals in Eq. 5 have closed-form solutions:

$$\int_{-\infty}^{+\infty} d\hat{E}_a \hat{E}_a \exp\left[\frac{(\hat{E}_a - E_a)^2}{2\sigma_g^2} - \beta \hat{E}_a\right] = \sqrt{2\pi}\sigma_g (E_a - \beta\sigma_g^2) \exp\left[\frac{\beta^2\sigma_g^2}{2} - \beta E_a\right] \quad (\text{S12})$$

and

$$\int_{-\infty}^{+\infty} d\hat{E}_a \exp \left[ \frac{(\hat{E}_a - E_a)^2}{2\sigma_g^2} - \beta \hat{E}_a \right] = \sqrt{2\pi} \sigma_g \exp \left[ \frac{\beta^2 \sigma_g^2}{2} - \beta E_a \right] \quad (\text{S13})$$

Introducing eqns. (S12) and (S13) into eqn. (S11) and simplifying gives

$$\langle \hat{E}_a \rangle_k = \frac{\int dE_a \tilde{\rho}(E_a) (E_a - \beta \sigma_g^2) e^{-\beta E_a}}{\int dE_a \tilde{\rho}(E_a) e^{-\beta E_a}} \quad (\text{S14})$$

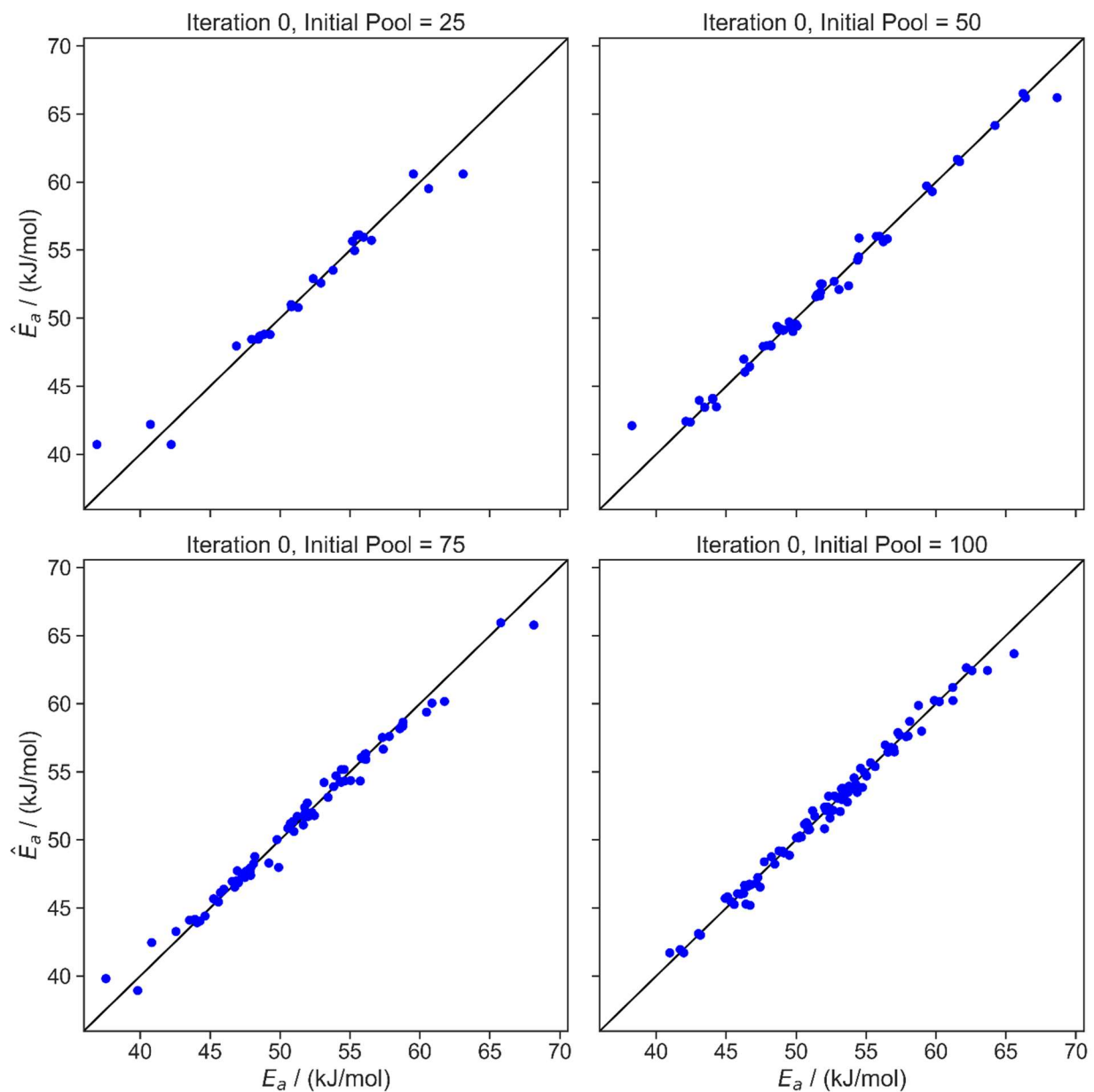
From eqn. (S8), it can be seen that

$$\langle \hat{E}_a \rangle_k = \langle E_a \rangle_k - \beta \sigma_g^2 . \quad (\text{S15})$$

Kernel regression errors ( $\sigma_g$ ) can be used to estimate the error in the kernel regression model predicted k-weighted activation barrier ( $\langle \hat{E}_a \rangle_k$ ) using eqn. (S9). We can estimate typical size of kernel regression errors using the training set error.

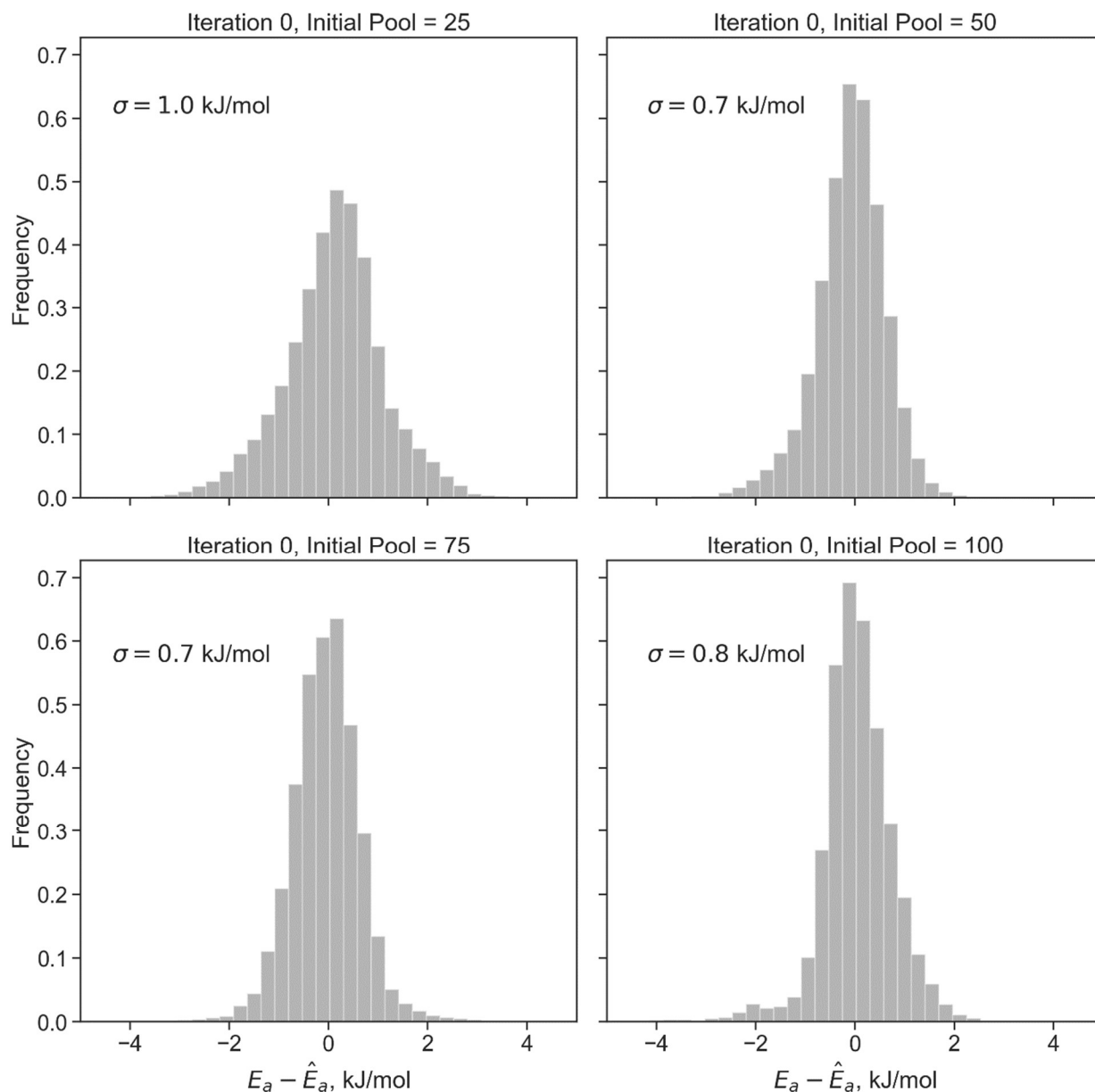
#### S4. Test set and training set statistics

The set of randomly sampled sites used to train the kernel regression model should sufficiently sample the main support of  $\tilde{\rho}(E_a)$  to properly normalize  $\tilde{\rho}(\hat{E}_a)$  for predicting kinetic properties. Once the main support of  $\tilde{\rho}(E_a)$  is sufficiently sampled, additional sites do not improve the normalization of  $\tilde{\rho}(\hat{E}_a)$  and require additional, costly structure optimizations. Fig. S1 shows the leave-one-out parity plot of the kernel regression plot trained on 25, 50, 75, and 100 randomly sampled sites.



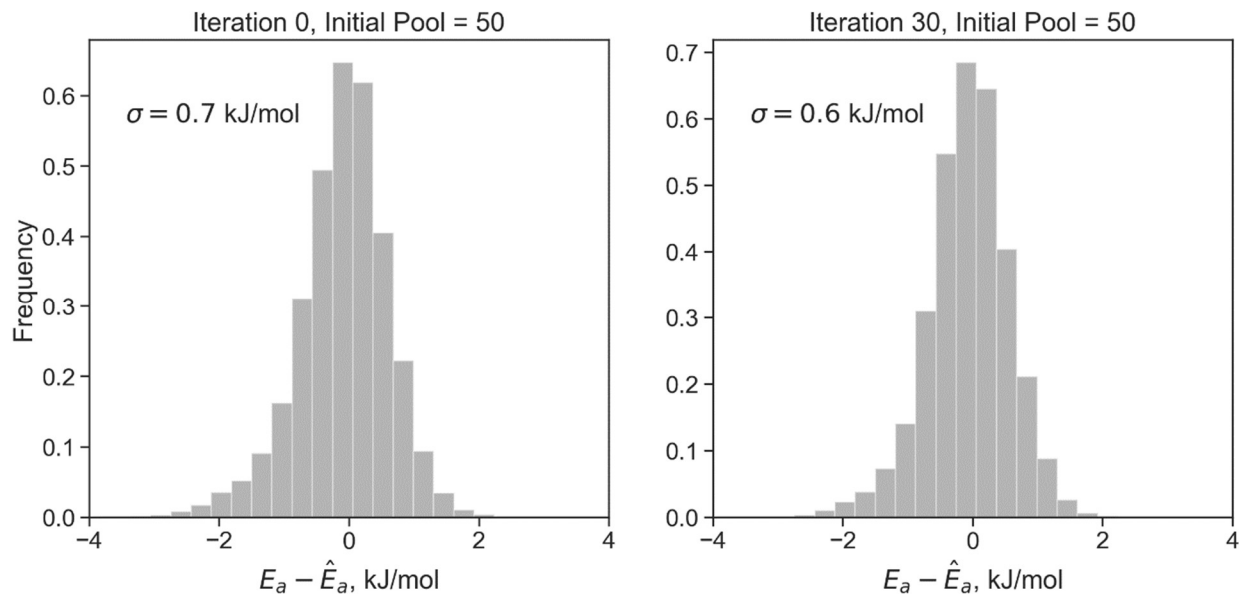
**Fig. S1:** Parity plot of kernel regression model trained on different initial pool sizes. An initial pool of 50 randomly selected sites samples the main support of  $\tilde{\rho}(E_a)$ .

Errors in the kernel regression model should be smaller than the width of  $\tilde{\rho}(E_a)$  to accurately importance sample  $\tilde{\rho}(\hat{E}_a)$ . Therefore, the initial pool should contain a set of sites with diverse local environments and activation energies to effectively train the kernel regression model. Residual distributions of all  $\sim 20,000$  sites are shown for the kernel regression model trained on 25, 50, 75, and 100 randomly sampled sites in Fig S2.



**Fig S2:** Kernel regression model residual distribution for all  $\sim 20,000$  sites with different initial pool sizes. For all initial pool sizes, the standard error is within 1.0 kJ/mol which is  $\sim 40$  times smaller than the range of  $\tilde{\rho}(\hat{E}_a)$ . The standard error does not decrease for initial pool sizes greater than 50.





**Fig S3:** Distribution of residuals for iterations 0 (left) and 30 (right) of the importance learning algorithm.

### S5. Number of samples required to estimate $\tilde{E}_a$ with the same precision $\bar{E}_a$

The  $\bar{E}_a$  estimator from the importance learning algorithm (eqn. 15) quickly converges to the correct site averaged activation energy because sites are sampled with weights  $\rho(\mathbf{x})k(\mathbf{x})$ . Alternatively, the  $\tilde{E}_a$  estimator randomly samples sites with weights  $\rho(\mathbf{x})$  and computes a ratio of exponential averages (eqn. 14). The reweighted estimator will require many more samples to converge to a precise estimate. In this section, the relative variance for the  $\tilde{E}_a$  estimator is derived and the number of samples required to estimate  $\tilde{E}_a$  with the same level of confidence as  $\bar{E}_a$  is computed.

From eqn. 14,  $\tilde{E}_a$  is computed by

$$\begin{aligned}\tilde{E}_a &= \frac{\sum_{i=1}^n k(\mathbf{x}_i)E_a(\mathbf{x}_i)}{\sum_i k(\mathbf{x}_i)} \\ &= \frac{\tilde{k}\tilde{E}_a}{\tilde{k}}\end{aligned}\tag{S16}$$

Both  $\tilde{k}\tilde{E}_a$  and  $\tilde{k}$  are random variables for a given sample size, so their ratio is also a random variable. Assuming  $\tilde{k}$  and  $\tilde{k}\tilde{E}_a$  are independent and uncorrelated, the sample variance of  $\tilde{E}_a$  can be approximated by:

$$\sigma_{\tilde{E}_a}^2 \approx \left(\frac{\partial \tilde{E}_a}{\partial \tilde{k}\tilde{E}_a}\right)^2 \sigma_{\tilde{k}\tilde{E}_a}^2 + \left(\frac{\partial \tilde{E}_a}{\partial \tilde{k}}\right)^2 \sigma_{\tilde{k}}^2\tag{S17}$$

Evaluating the derivatives and dividing by  $\tilde{E}_a$  yields the relative sample variance

$$\frac{\sigma_{\tilde{E}_a}^2}{\tilde{E}_a^2} = \frac{\sigma_{\tilde{k}\tilde{E}_a}^2}{(\tilde{k}\tilde{E}_a)^2} + \frac{\sigma_{\tilde{k}}^2}{\tilde{k}^2}\tag{S18}$$

The relative sample variance can be related to the relative variance by the central limit theorem:<sup>7</sup>

$$\frac{\sigma_{\tilde{k}E_a}^2}{(\tilde{k}\bar{E}_a)^2} + \frac{\sigma_{\tilde{k}}^2}{\tilde{k}^2} = \frac{1}{N} \left[ \frac{\sigma_{kE_a}^2}{\langle kE_a \rangle^2} + \frac{\sigma_k^2}{\langle k \rangle^2} \right]_{\rho(\mathbf{x})} \equiv \frac{\sigma_{\langle E_a \rangle_k}^2}{\langle E_a \rangle_k^2} \Big|_{\rho(\mathbf{x})} \quad (\text{S19})$$

where  $N$  is the number of samples. The right most equality with subscript  $\rho(\mathbf{x})$  indicates that (S19) estimates the relative variance in the  $\bar{E}_a$  estimate as computed with a sample from  $\rho(\mathbf{x})$ . The number of random samples required to match the uncertainty of the  $\bar{E}_a$  estimator from the importance learning algorithm is found by equating the relative uncertainties of the two estimators:

$$\frac{\sigma_{\langle E_a \rangle_k}^2}{\langle E_a \rangle_k^2} \Big|_{k(\mathbf{x})\rho(\mathbf{x})} = \frac{\sigma_{\langle E_a \rangle_k}^2}{\langle E_a \rangle_k^2} \Big|_{\rho(\mathbf{x})} \quad (\text{S20})$$

Inserting eqn. (S19) in the right hand side of (S20) and solving for  $N$  yields

$$N = \left[ \frac{\sigma_{\langle E_a \rangle_k}^2}{\langle E_a \rangle_k^2} \right]_{k(\mathbf{x})\rho(\mathbf{x})}^{-1} \left[ \frac{\sigma_{kE_a}^2}{\langle kE_a \rangle^2} + \frac{\sigma_k^2}{\langle k \rangle^2} \right]_{\rho(\mathbf{x})} \quad (\text{S21})$$

The relative uncertainty in the  $\bar{E}_a$  estimator is  $(0.75 \text{ kJ/mol}) / (40.5 \text{ kJ/mol}) = 1.85 \%$ . Since  $\tilde{\rho}(E_a)$  can be precisely calculated for our simple model,  $\sigma_{kE_a}^2 / \langle kE_a \rangle^2$  and  $\sigma_k^2 / \langle k \rangle^2$  can be computed exactly. Evaluating eqn. (S21) gives

$$N = (0.0185)^{-2} \times (28.1 + 41.2) \approx 200,000 \quad (\text{S22})$$

Therefore, the reweighting estimator  $\tilde{E}_a$  requires about 200,000 sites for the same level of confidence that the importance learning estimator  $\bar{E}_a$  achieved with less than 100 sites.

## Additional References

1. A. Fong, C. Vandervelden, S. L. Scott and B. Peters, *ACS Catal.*, 2018, 1728–1733.
2. A. Fong, Y. Yuan, S. L. Ivry, S. L. Scott and B. Peters, *ACS Catal.*, 2015, **5**, 3360-3374.
3. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16 Rev. B.01, 2016
4. J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.
5. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
6. A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571-2577.
7. J. D. Logan, *Applied mathematics*, Wiley-Interscience, Hoboken, N.J., 3rd edn., 2006.