Supporting Information for

"Machine Learning Quantum Reaction Rate Constants"

Evan Komp[*1] and Stéphanie Valleau,[*2]

[*]Department of Chemical Engineering, University of Washington, Seattle, Washington 98115, United States

Table of Contents

[1] email: evankomp@uw.edu

[2] email: valleau@uw.edu

I. DATABASE GENERATION

The equations for the potentials employed to generate the database are defined below in Table S1. For the single barriers the exact expressions for the transmission coefficient $T(E)$ were used.[1,2] The first asymmetry parameter α for a single barrier was defined as the difference in energy between products and reactants. The second asymmetry parameter α_2 was defined as $\alpha_2 = \frac{V_1 - V_2}{V_1 + V_2} + \frac{w_1 - w_2}{w_1 + w_2}$, with V_1, V_2 the height of the barriers and w_1, w_2 their width.

For double barriers, the transmission coefficient was evaluated as in equation (2) of the main text[3], with $x \in [-L, L]$ and $L = 20$ [au]. The evaluation of the transmission coefficient was carried out using scipy's `integrate.solve_ivp` function to solve the initial value problem[4] with the implicit Runge-Kutta method of order 5 as defined by the '`Radau`' option. The range of energy values over which the transmission coefficient was evaluated was a set of 300 points in the interval $E_{values} = \left[\frac{E_{max}}{1000}; 2E_{max}\right]$, with $E_{max} = 2 \cdot \max[V(x)]$. The maximum value of the potential energy was evaluated over 1000 points in the range $-L < x < L$.

Table S1: Equations for the one-dimensional potential energy barriers used to build the database.

Barrier	Potential energy	Parameters
Single rectangular	$\mathrm{V_R}(x) = \begin{cases} \mathrm{V_1} & 0.0 < x < \mathrm{w_1} \\ 0.0 & x < 0 \ \vee \ x > \mathrm{w_1} \end{cases}$	$\mathrm{V_1}$ – barrier height $\mathrm{w_1}$ – barrier width $\alpha_1 = 0.0$ $\alpha_2 = 0.0$ $s = 0.0$
Single symmetric Eckart	$\mathrm{V_{SE}}(x) = \frac{\mathrm{V_1}}{\cosh^2(\pi x / \mathrm{w_1})}$	$\mathrm{V_1}$ – barrier height $\mathrm{w_1}$ – barrier width $\alpha_1 = 0.0$ $\alpha_2 = 0.0$ $s = \frac{\mathrm{V}(x^*) - \mathrm{V}\left(x^* + \frac{w_1}{2} - \Delta x\right)}{w_1/2}$ $x^* \mid \mathrm{V}(x^*) = \max(\mathrm{V}(x))$ $\Delta x = 0.025$ [au]
Single asymmetric Eckart	$\mathrm{V_{AE}}(x) = \frac{\mathrm{V_1}(1-\alpha)}{1 + \mathrm{e}^{-2\pi x / w_1}} + \frac{\mathrm{V_1}\left(1 + \sqrt{\alpha}\right)^2}{4\cosh^2 \pi x / w_1}$	$\mathrm{V_1}$ – barrier height $\mathrm{w_1}$ – barrier width $\alpha_1 = V_{react} - V_{prod}$ $\alpha_2 = 0.0$ $s = \frac{\lvert \mathrm{V}(x^*) - \mathrm{V}(x_+) \rvert + \lvert \mathrm{V}(x^*) - \mathrm{V}(x_-) \rvert}{w_1}$ $x^* \mid \mathrm{V}(x^*) = \max(\mathrm{V}(x))$ $x_\pm = x^* \pm \frac{w_1}{2} \mp \Delta x \ ; \ \Delta x = 0.025$ [au]

Double rectangular	$\mathrm{V_{DR}}(x) = \begin{cases} \mathrm{V_1} & 0.0 < x < \mathrm{w_1} \\ \mathrm{V_2} & \mathrm{w_1} + \Delta x_{12} < x < \mathrm{w_1} + \Delta x_{12} + w_2 \\ 0.0 & \text{elsewhere} \end{cases}$	V_i – barrier *i* height w_i – barrier *i* width $d = \Delta x_{12} + (w_1 + w_2)/2$ $\alpha_1 = 0$ $\alpha_2 = \frac{V_1 - V_2}{V_1 + V_2} + \frac{w_1 - w_2}{w_1 + w_2}$ $s = 0.0$
Double gaussian	$\mathrm{V_{DG}}(x) = \mathrm{V_1 e}^{\frac{-(x-x_1)^2}{2\sigma_1^2}} + \mathrm{V_2 e}^{\frac{-(x-x_2)^2}{2\sigma_2^2}}$ $\sigma_1 = \mathrm{w_1}/3\,; \quad \sigma_2 = \mathrm{w_2}/3$ $x_1 = 0.0\,; \quad x_2 = 3 \cdot (\sigma_1 + \sigma_2) + d$	V_i – barrier *i* height w_i – barrier *i* width *d* – distance between barriers $\alpha_1 = 0$ $\alpha_2 = \frac{V_1 - V_2}{V_1 + V_2} + \frac{w_1 - w_2}{w_1 + w_2}$ $s = (s_1 + s_2)/2$ $s_i = \frac{\mathrm{V}(x_i^*) - \mathrm{V}\left(x_i^* + \frac{w_i}{2} - \Delta x\right)}{\frac{w_i}{2}}$ $x_i^* \mid V(x_i^*) = \max\big(V(x)\big)\ ; i = 1{,}2$ $\Delta x = 0.025\ [au]$
Peskin barrier	$\mathrm{V_{Pesk}}\,(x) = \mathrm{V}'\left(\frac{1}{\cosh^2(x)} - \frac{1}{\cosh^2(\gamma x)}\right)$	V' – constant proportional to barrier height γ – constant proportional to barrier width $\mathrm{w_1} = w_2 = \lvert x^* \ - \mathrm{x}_{min} \rvert$ barrier width $\alpha_1 = 0.0$ $\alpha_2 = 0.0$ $s = (s_1 + s_2)/2 s_i = \frac{\mathrm{V}(x_i^*) - \mathrm{V}\left(x_i^* + \frac{w_i}{2} - \Delta x\right)}{w_i/2}$; difference taken with respect to $x = 0.0$ when $V_{max} > 0 \wedge V_{min} > 0$ $x^* \mid V(x^*) = \max(V(x))$

II. DATASET

A visualization of the distribution of the training and testing dataset with respect to its input features and output label is shown in Figures S1a-c. The first plot (a) displays scattering of pairwise feature distributions with kernel density estimate (KDE) diagonals for the entire dataset. The following plots (b-c) show the pairwise distribution of subsets of features and labels for the test and train sets through KDE. Figure S2 maps the correlation between variables through the Pearson correlation coefficient.

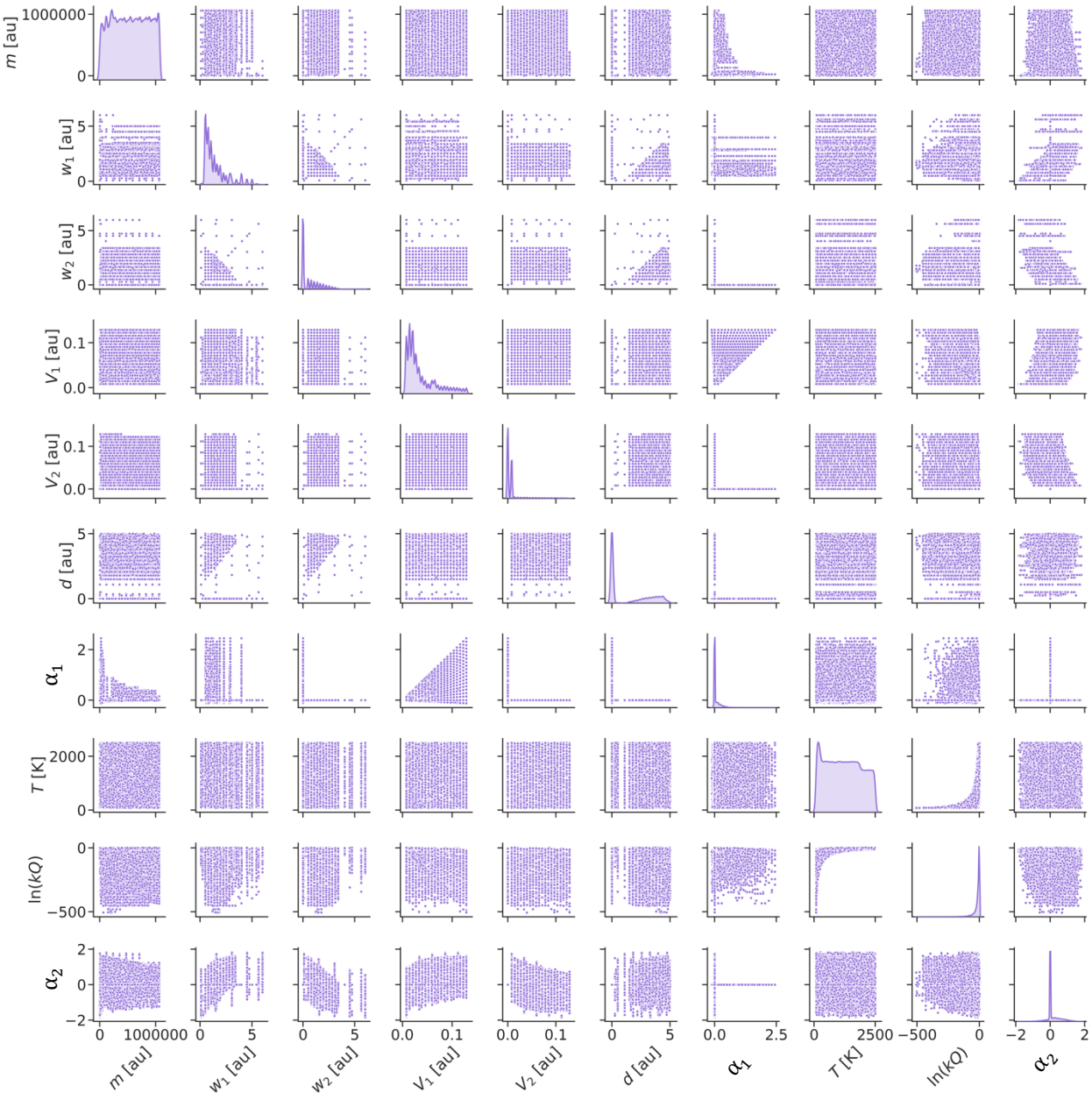

Figure S1a: Database pair-wise distribution for all input features and for the output label. Points are datum. Diagonals are gaussian KDE of single features.

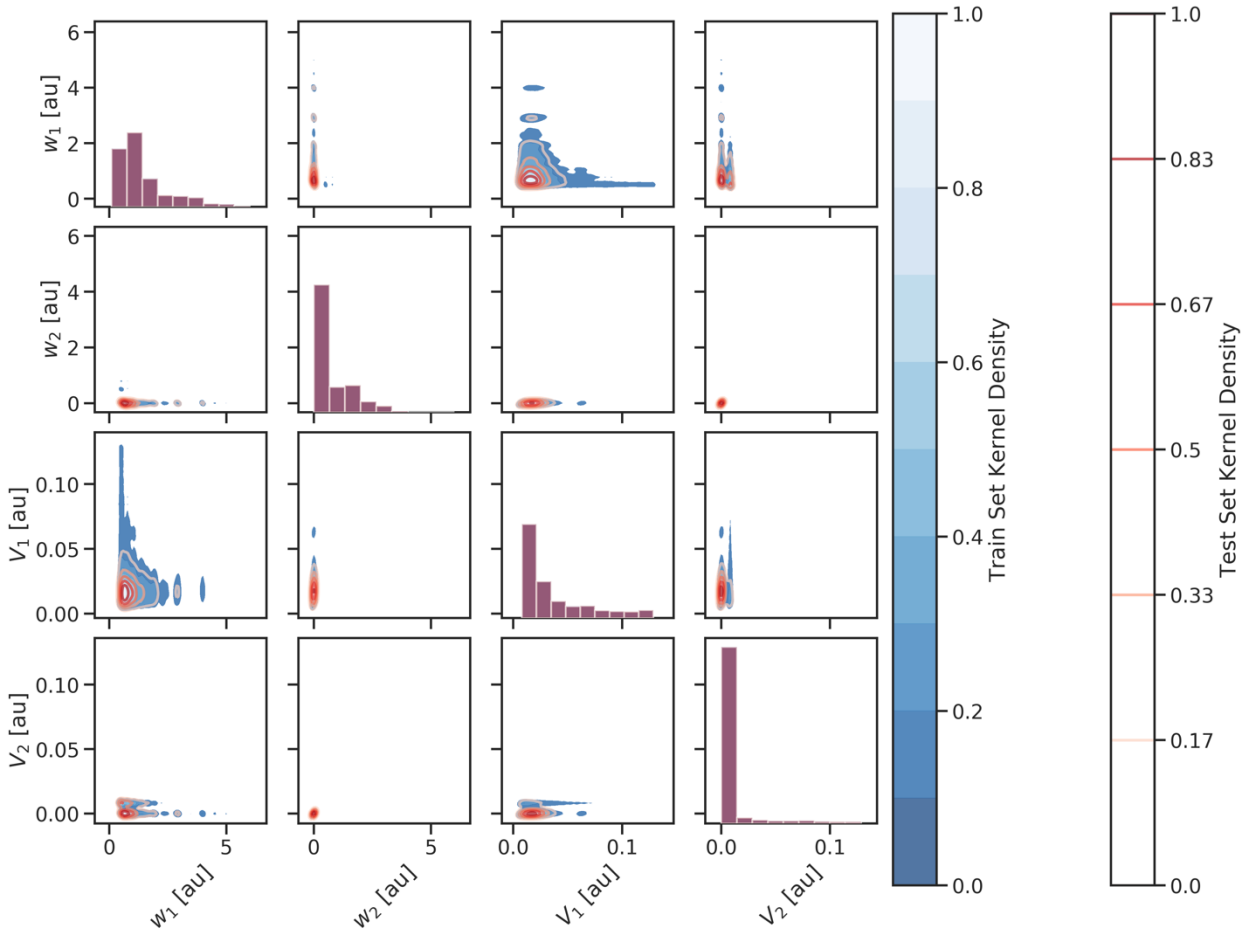

Figure S1b: Pair-wise distribution of a subset of the database features related to the barrier for the train and test sets. Blue surfaces are train set KDE, while red lines are test set KDE. KDE is normalized to 1.0. A KDE of 1.0 then indicates the region of two-feature space with the highest density of data.

When looking at the diagonal distribution of weights and heights in Figure S1b, we see a large number of points in the first bin. This comes from the fact that for single barriers the distance between barriers, the second barrier width and height are all equal to zero.

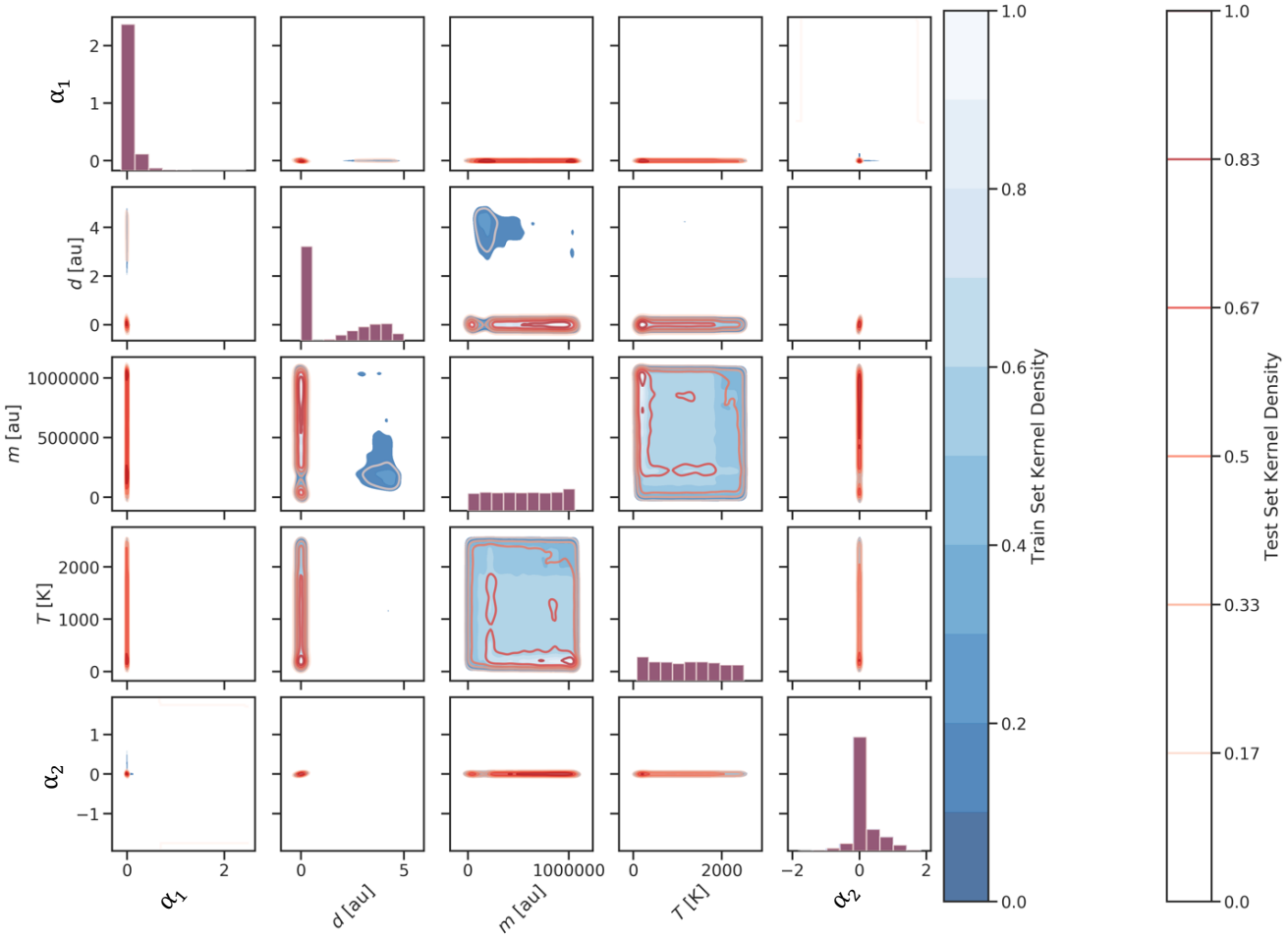

Figure S1c: Pair-wise distribution of a subset of the database features for the train and test sets. Blue surfaces are train set KDE, while red lines are test set KDE. KDE is normalized to 1.0. A KDE of 1.0 then indicates the region of two-feature space with the highest density of data.

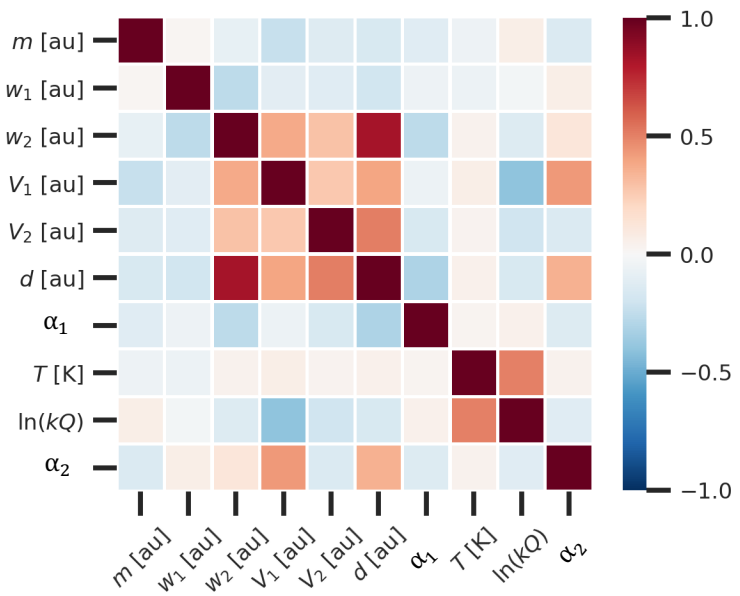

Figure S2: Correlation matrix of the Pearson correlation coefficient for input features of the database. Values close to 1.0 indicate a strong positive correlation between variables, while -1.0 indicates a strong negative correlation. Features are perfectly correlated to themselves, as can be seen by the dark red diagonal. Most features have a small correlation, close to zero which indicates little bias. One distinct correlation is between barrier distance and second barrier width. This is due to the fact that for single barriers d and w_2 are zero while for double barriers both are non-zero. We also note a positive correlation between temperature and rate, which comes from the physics of reactivity.

III. GRID SEARCH & HYPERPARAMETER TUNING

As described in the main text, four subsequent grid searches were carried out. The hyperparameter search spaces for each are shown in Table S2-5, and optimal hyperparameters are shown in Table S6. Figure S 3 indicates how the logarithm of the mean absolute error, MSE, averaged over all runs, changes with the number of neurons for each single hidden layer. Figure S 4 shows how the loss changes with the learning rate. Figure S5 shows validation loss and moving average validation loss during training for a range of batch sizes, highlighting the optima.

A. First grid search: neurons, hidden layers, batch size and epochs

For the first grid search we searched over number of hidden layers, number of neurons per layer, and batch size. The range of parameters is shown in Table S2 below and the results in the following Figure S 3. We found that the optimal model was of neuron configuration (64, 24, 24).

Table S2: First grid search space values for number of hidden layers, number of neurons per layer, batch size and epochs.

Hyperparameter	**Search space**
number of hidden layers	1, 2, 3
number of neurons in hidden layer	1, 6, 12, 24, 32, 64, 128
batch size	32, 64, 128
Fixed hyperparameters	**Default values**
hidden layer activation function	softsign
output activation function	tanh
epochs	300
optimizer	Adam
- learning rate	0.001
- beta 1	0.9
- beta 2	0.9999
weight initialization	random uniform
- range	[0, 1]
loss function	mean squared error (MSE)
Optimal hyperparameters	**Values**
neuron configuration	(64, 24, 24)
batch size	64

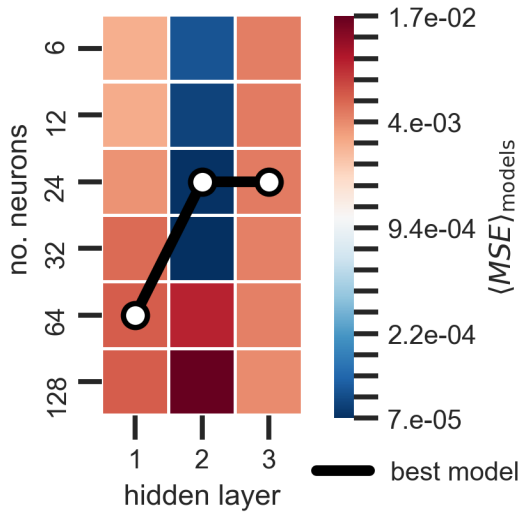

Figure S 3: Average model validation mean square error (MSE) associated with number of neurons in each of three hidden layers. The color scale corresponds to the average of the MSE over all models with three hidden layers trained for 300 epochs with batch size 64. The black line shows the neuron configuration of the optimal model.

B. Second grid search: activation function

For the second grid search, the choice of activation functions applied to each neuron in the hidden and output layers was explored. The activation functions tested are seen below in Table S3. The optimum configuration was found to be softsign, tanh for hidden and output layers, respectively.

Table S3: Second grid search space values for activation function in both hidden and output layers.

Hyperparameter	**Search space**
activation function	softmax, softplus, softsign, relu, tanh, sigmoid, hard sigmoid linear
Fixed hyperparameters	**Default values**
neuron configurations	(64, 24, 64)
epochs	300
batch size	32
optimizer	Adam
- learning rate	0.001
- beta 1	0.9
- beta 2	0.9999
weight initialization	random uniform
- range	[0, 1]
loss function	mean squared error (MSE)
Optimal hyperparameters	**Values**
hidden layer activation	softsign
output layer activation	tanh

C. Third grid search: learning rate

The third grid search optimized the value of learning rate for the Adam optimizer. The range of values explored were on the logarithmic scale to best search hyperparameter space and are shown below in Table S4. This was conducted on a randomly selected portion of 30% of the training set as well as on 100% of the training set. In both cases, the best value was 0.0005. The results are also shown in Figure S 4.

Table S4: Third grid search space values for the learning rate.

Hyperparameter	**Search space**
learning rate	1E-5, 5E-5, 1E-4, 5E-4,1E-3, 5E-3,1E-2, 5E-2, 1E-1, 2E-1
Fixed Hyperparameters	**Default values**
neuron configuration	(64, 24, 24)
epochs	300
batch size	64
hidden layer activation function	softsign
output activation function	tanh
optimizer	Adam
- beta 1	0.9
- beta 2	0.9999
weight initialization	random uniform
- range	[0, 1]
loss function	mean squared error (MSE)
Optimal hyperparameters	**Values**
learning rate	0.0005

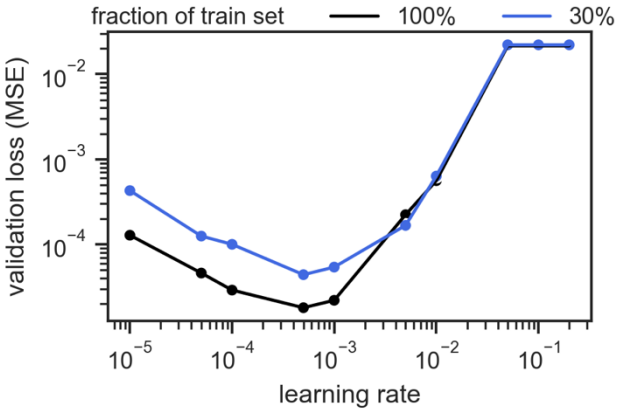

Figure S 4: Validation loss as a function of learning rate in the third grid search. The blue and black lines correspond to grid search on 30% and 100% of the training dataset. Both searches indicate that 0.0005 is the optimal learning rate value.

D. Fourth grid search: batch size for optimal model

A final grid search on the batch size was conducted to evaluate how fluctuations on the validation loss related to mini batch training batch sizes. Details of the search are described in Table S5. A moving average of the validation loss was used as a metric, to average the impact of the fluctuations. It was found that 128, and 256 performed best in terms of reduction in fluctuation and final validation loss magnitude and given that 128 had a lower final rolling average loss it was chosen as the best batch size. This result is shown in Figure S5.

Table S5: Fourth grid search space for batch size.

Hyperparameter	**Search space**
batch size	32, 64, 128, 256, 512, 1024, 2048, 4096, 8192
Fixed Hyperparameters	**Default values**
neuron configuration	(64, 24, 24)
epochs	300
hidden layer activation function	softsign
output activation function	tanh
optimizer	Adam
- learning rate	0.0005
- beta 1	0.9
- beta 2	0.9999
weight initialization	random uniform
- range	[0, 1]
loss function	mean squared error (MSE)
Optimal hyperparameters	**Values**
batch size	128

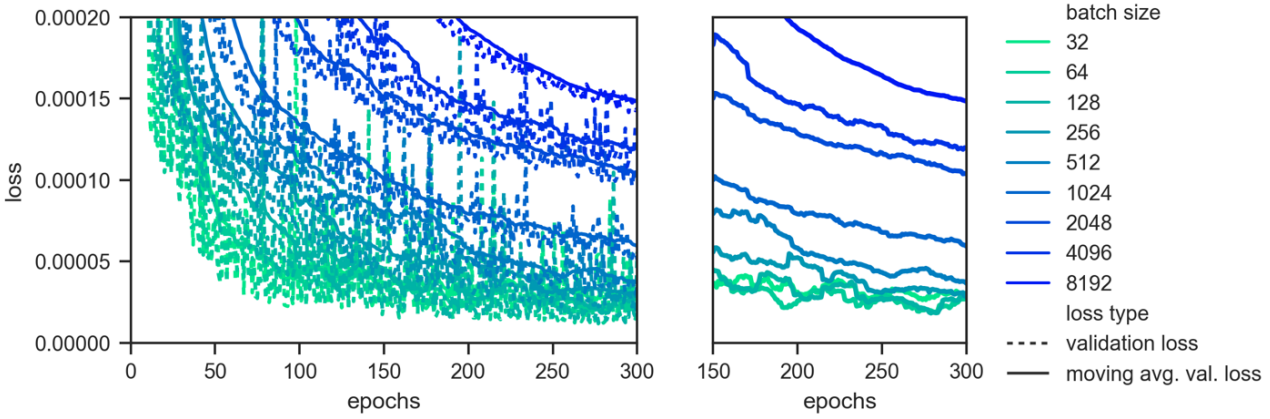

Figure S5: Left panel) Validation loss and moving average of validation loss over 20 points for models trained with a range of batch sizes as a function of epochs. Batch size is seen to impact both the fluctuations in the validation loss, as well as the final training and validation loss value. Right panel) Moving average validation loss for the last 100 epochs of training. Batch sizes 128, and 256 perform best.

E. Optimal hyperparameters

Table S6: Optimal hyperparameters identified from grid searches for the final training.

Hyperparameter	**Value**
neuron configuration	(64, 24, 24)
epochs	300
batch size	128
hidden layer activation function	softsign
output activation function	tanh
optimizer	Adam
- learning rate	0.0005
- beta 1	0.9
- beta 2	0.9999
weight initialization	random uniform
- range	[0, 1]
loss function	mean squared error (MSE)
Performance	
training loss	4.02E-6
test loss	6.18E-6
test MAE [log(1/ps)]	0.18 (0.60)

IV. INPUT FEATURES FOR RATE CONSTANTS PREDICTED IN FIGURE 6 OF MAIN TEXT

Table S7 contains the values of the input features provided to the trained DNN to compute reaction rate constants for the reactions shown in Figure 6 of the main text. The input features were obtained by fitting existing minimum energy paths to asymmetric Eckart barriers.

Table S7: Input feature values used to make predictions on the rate product for asymmetric reactions outside of the testing set. Reaction 1 is the Menshutkin reaction of pyridine with methyl bromide in the gas phase.[5] Reaction 2 is the reaction of formalcyanohydrin with hydrogen sulfide in the liquid phase,[6] and Reaction 3 is the reaction of F with HCl.[7] Reaction 3a uses the activation energy obtained from the Eckart fit of the minimum energy path, while Reaction 3b uses a higher level of theory activation energy.[7]

Feature	Reaction 1	Reaction 2	Reaction 3a	Reaction 3b
mass [au]	317257.3	226395.5	101090.8	101090.8
V_1 [au]	0.04223043	0.03107522	0.01149779	0.002533826
V_2 [au]	0.0	0.0	0.0	0.0
slope [au]	0.008847832	0.01289512	0.03025426	0.03025426
w_1 [au]	4.314845	4.0	2.303736	2.303736
w_2 [au]	0.0	0.0	0.0	0.0
d [au]	0.0	0.0	0.0	0.0
α_1	-0.03879398	0.003222610	0.05662957	0.05662957
α_2	0.0	0.0	0.0	0.0

REFERENCES

(1) Eckart, C. The Penetration of a Potential Barrier by Electrons. *Phys. Rev.* **1930**, *35* (11), 1303–1309. https://doi.org/10.1103/PhysRev.35.1303.

(2) Ahmed, Z. Tunneling through a One-Dimensional Potential Barrier. *Phys. Rev. A* **1993**, *47* (6), 4761–4767. https://doi.org/10.1103/PhysRevA.47.4761.

(3) Mandrà, S.; Valleau, S.; Ceotto, M. Deep Nuclear Resonant Tunneling Thermal Rate Constant Calculations. *Int. J. Quantum Chem.* **2013**, *113* (12), 1722–1734. https://doi.org/10.1002/qua.24395.

(4) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. https://doi.org/https://doi.org/10.1038/s41592-019-0686-2.

(5) Castejon, H.; Wiberg, K. B. Solvent Effects on Methyl Transfer Reactions. 1. The Menshutkin Reaction. *J. Am. Chem. Soc.* **1999**, *121* (10), 2139–2146. https://doi.org/10.1021/ja983736t.

(6) Valleau, S.; Martínez, T. J. Reaction Dynamics of Cyanohydrins with Hydrosulfide in Water. *J. Phys. Chem. A* **2019**, *123* (33), 7210–7217. https://doi.org/10.1021/acs.jpca.9b05735.

(7) Aoto, Y. A.; Köhn, A. Revisiting the F + HCl → HF + Cl Reaction Using a Multireference Coupled-Cluster Method. *Phys. Chem. Chem. Phys.* **2016**, *18* (44), 30241–30253. https://doi.org/10.1039/c6cp05782a.